



THE UNIVERSITY *of* EDINBURGH

## Edinburgh Research Explorer

### **BacHBerry: BACterial Hosts for production of Bioactive phenolics from bERRY fruits**

**Citation for published version:**

Dudnik, A, Almeida, AF, Andrade, R, Bañados, P, Barbay, D, Bassard, J-E, Benkoulouche, M, Bott, M, Braga, A, Breitel, D, Brennan, R, Bulteau, L, Chanforan, C, Costa, I, Costa, RS, Doostmohammadi, M, Faria, N, Feng, C, Fernandes, A, Ferreira, P, Ferro, R, Foito, A, Freitag, S, Garcia, G, Gaspar, P, Godinho-Pereira, J, Hamberger, B, Hartmann, A, Heider, H, Jardim, C, Julien-Laferriere, A, Kallscheuer, N, Kerbe, W, Kuipers, OP, Li, S, Love, N, Marchetti-Spaccamela, A, Marienhagen, J, Martin, C, Mary, A, Mazurek, V, Meinhart, C, Méndez Sevillano, D, Menezes, R, Naesby, M, Nørholm, MHH, Okkels, FT, Oliveira, J, Ottens, M, Parrot, D, Pei, L, Rocha, I, Rosado-Ramos, R, Rousseau, C, Sagot, M-F, Nunes dos Santos, C, Schmidt, M, Shelenga, T, Shepherd, L, Silva, AR, da Silva, MH, Simon, O, Stahlhut, SG, Solopova, A, Sorokin, A, Stewart, D, Stougie, L, Su, S, Thole, V, Tikhonova, O, Trick, M, Vain, P, Veríssimo, A, Vila-Santa, A, Vinga, S, Vogt, M, Wang, L, Wang, L, Wei, W, Youssef, S, Rute Neves, A & Forster, J 2018, 'BacHBerry: BACterial Hosts for production of Bioactive phenolics from bERRY fruits', *Phytochemistry Reviews*, vol. 17, no. 2, pp. 291-326. <https://doi.org/10.1007/s11101-017-9532-2>

**Digital Object Identifier (DOI):**

[10.1007/s11101-017-9532-2](https://doi.org/10.1007/s11101-017-9532-2)

**Link:**

[Link to publication record in Edinburgh Research Explorer](#)

**Document Version:**

Peer reviewed version

**Published In:**

Phytochemistry Reviews

**General rights**

Copyright for the publications made accessible via the Edinburgh Research Explorer is retained by the author(s) and / or other copyright owners and it is a condition of accessing these publications that users recognise and abide by the legal requirements associated with these rights.

**Take down policy**

The University of Edinburgh has made every reasonable effort to ensure that Edinburgh Research Explorer content complies with UK legislation. If you believe that the public display of this file breaches copyright please contact [openaccess@ed.ac.uk](mailto:openaccess@ed.ac.uk) providing details, and we will remove access to the work immediately and investigate your claim.



[Click here to view linked References](#)

1    **TITLE**

2    BacHBerry: BACterial Hosts for production of Bioactive phenolics from bERRY fruits

3    **RUNNING TITLE**

4    BacHBerry

5

6    **AUTHORS**

7    The list of contributing authors is given on p. 2.

8

9    \*Corresponding author:

10    Alexey Dudnik

11    Email: [adud@biosustain.dtu.dk](mailto:adud@biosustain.dtu.dk)

12    Phone: + 45 93 51 11 01

13    Fax: +45 45 25 80 01

14

15

16    Keywords: berries; bioprospecting; microbial cell factories; polyphenols; sustainable production.

17

18    **Footnote**

19    This article is written by the BacHBerry consortium ([www.bachberry.eu](http://www.bachberry.eu)) and represents the collective effort  
20    of all participating institutions. The authors are therefore listed in alphabetical order.

21 **The list of authors**

22 Alexey Dudnik<sup>1,\*</sup>, A. Filipa Almeida<sup>15,20</sup>, Ricardo Andrade<sup>2</sup>, Barbara Avila<sup>3</sup>, Pilar Bañados<sup>3</sup>, Diane Barbay<sup>4</sup>,  
23 Jean-Etienne Bassard<sup>5</sup>, Mounir Benkoulouche<sup>4</sup>, Michael Bott<sup>6</sup>, Adelaide Braga<sup>7, 21</sup>, Dario Breitel<sup>12</sup>, Rex  
24 Brennan<sup>10</sup>, Laurent Bulteau<sup>2</sup>, Celine Chanforan<sup>8</sup>, Inês Costa<sup>15,20</sup>, Rafael S. Costa<sup>9</sup>, Mahdi  
25 Doostmohammadi<sup>9,22</sup>, Nuno Faria<sup>7, 21</sup>, Chengyong Feng<sup>17</sup>, Armando Fernandes<sup>9</sup>, Patricia Ferreira<sup>7, 21</sup>, Roberto  
26 Ferro<sup>1</sup>, Alexandre Foito<sup>10</sup>, Sabine Freitag<sup>10</sup>, Gonçalo Garcia<sup>15,20</sup>, Paula Gaspar<sup>1</sup>, Joana Godinho-Pereira<sup>15,20</sup>,  
27 Björn Hamberger<sup>5</sup>, András Hartmann<sup>9</sup>, Harald Heider<sup>4</sup>, Carolina Jardim<sup>15,20</sup>, Alice Julien-Laferriere<sup>2</sup>, Nicolai  
28 Kallscheuer<sup>6</sup>, Wolfgang Kerbe<sup>13</sup>, Oscar P. Kuipers<sup>11</sup>, Shanshan Li<sup>17</sup>, Nicola Love<sup>12</sup>, Alberto Marchetti-  
29 Spaccamela<sup>2</sup>, Jan Marienhagen<sup>6</sup>, Cathie Martin<sup>12</sup>, Arnaud Mary<sup>2</sup>, Vincent Mazurek<sup>8</sup>, Camillo Meinhardt<sup>13</sup>,  
30 David Méndez Sevillano<sup>14</sup>, Regina Menezes<sup>15,20</sup>, Michael Naesby<sup>4</sup>, Morten H.H. Nørholm<sup>1</sup>, Finn T. Okkels<sup>8</sup>,  
31 Joana Oliveira<sup>7, 21</sup>, Marcel Ottens<sup>14</sup>, Delphine Parrot<sup>2</sup>, Lei Pei<sup>13</sup>, Isabel Rocha<sup>7, 21</sup>, Rita Rosado-Ramos<sup>15,20</sup>,  
32 Caroline Rousseau<sup>4</sup>, Marie-France Sagot<sup>2</sup>, Claudia Nunes dos Santos<sup>15,20</sup>, Markus Schmidt<sup>13</sup>, Tatiana  
33 Shelenga<sup>16</sup>, Louise Shepherd<sup>10</sup>, Ana Rita Silva<sup>7, 21</sup>, Marcelo Henriques da Silva<sup>14</sup>, Olivier Simon<sup>8</sup>, Steen  
34 Gustav Stahlhut<sup>1</sup>, Ana Solopova<sup>11</sup>, Artem Sorokin<sup>16</sup>, Derek Stewart<sup>10,19</sup>, Leen Stougie<sup>2</sup>, Shang Su<sup>17</sup>, Vera  
35 Thole<sup>12</sup>, Olga Tikhonova<sup>16</sup>, Martin Trick<sup>12</sup>, Philippe Vain<sup>12</sup>, André Veríssimo<sup>9</sup>, Ana Vila-Santa<sup>9</sup>, Susana  
36 Vinga<sup>9</sup>, Michael Vogt<sup>6</sup>, Liangsheng Wang<sup>17</sup>, Lijin Wang<sup>17</sup>, Wei Wei<sup>17</sup>, Sandra Youssef<sup>13</sup>, Ana Rute Neves<sup>18</sup>  
37 & Jochen Forster<sup>1</sup>

38

39 <sup>1</sup> The Novo Nordisk Foundation Center for Biosustainability, Technical University of Denmark, Kemitorvet,  
40 Building 220, 2800 Kgs. Lyngby, Denmark

41 <sup>2</sup> Institut National de Recherche en Informatique et Automatique, 43 Boulevard du 11 Novembre 1918,  
42 69622 Villeurbanne Cedex, France

43 <sup>3</sup> Facultad De Agronomía e Ingeniería Forestal, Pontificia Universidad Católica de Chile, Av. Vicuña  
44 Mackenna Ote, 4860 Macul, Chile

45 <sup>4</sup> Evolva, Duggingenstrasse 23, 4053 Reinach, Switzerland

46 <sup>5</sup> Department of Plant and Environmental Science, University of Copenhagen, 1871 Frederiksberg, Denmark

47 <sup>6</sup> Forschungszentrum Jülich GmbH, Institut für Bio- und Geowissenschaften, IBG-1: Biotechnologie,  
48 Wilhelm-Johnen-Straße, 52428 Jülich, Germany

49 <sup>7</sup> Biotempo, Avepark – Edif. Spinpark, Zona Industrial da Gandra, Barco, 4805-017 Guimarães, Portugal

50 <sup>8</sup> Chr. Hansen Natural Colors A/S, Agern Alle 24, 2970 Hørsholm, Denmark

51 <sup>9</sup> IDMEC, Instituto Superior Técnico, Universidade de Lisboa, Av. Rovisco Pais 1, 1049-001 Lisboa,  
52 Portugal

53 <sup>10</sup> The James Hutton Institute, Invergowrie, DD2 5DA Dundee, Scotland, United Kingdom

54 <sup>11</sup> Groningen Biomolecular Sciences and Biotechnology Institute, Department of Molecular Genetics,  
55 University of Groningen, Linnaeusborg, Nijenborgh 7, 9747 AG Groningen, The Netherlands

56 <sup>12</sup> John Innes Centre, Norwich Research Park, NR4 7UH Norwich, United Kingdom

57 <sup>13</sup> Biofaction KG, Kundmannsgasse 39/12, 1030 Vienna, Austria

58 <sup>14</sup> Department of Biotechnology, Delft University of Technology, van der Maasweg 9, 2629 HZ Delft, The  
59 Netherlands

60 <sup>15</sup> Instituto de Biologia Experimental e Tecnológica, Av. República, Qta. do Marquês, 2780-157 Oeiras,  
61 Portugal

62 <sup>16</sup> Fruit Crops Genetic Resources Department, N. I. Vavilov Research Institute of Plant Industry, B.  
63 Morskaya Street 42-44, 190000 St. Petersburg, Russia

64 <sup>17</sup> Institute of Botany, The Chinese Academy of Sciences, 20 Nanxincun, Xiangshan, 100093 Beijing, China

65 <sup>18</sup> Chr. Hansen Holding A/S, Bøge Alle 10-12, 2970 Hørsholm, Denmark

66 <sup>19</sup> Institute of Mechanical, Process and Energy Engineering, School of Engineering and Physical Sciences,  
67 Heriot Watt University, Edinburgh, Scotland, UK

68 <sup>20</sup> Instituto de Tecnologia Química e Biológica António Xavier, Universidade Nova de Lisboa, Av. da  
69 República, 2780-157 Oeiras, Portugal.

70 <sup>21</sup> Centre of Biological Engineering, University of Minho, Campus de Gualtar 4710-057, Braga, Portugal

71 <sup>22</sup> School of Mathematics, University of Edinburgh, EH9 3FD Edinburgh, Scotland, United Kingdom

## 72    **Abstract**

73    BACterial Hosts for production of Bioactive phenolics from bERRY fruits (BacHBerry) was a three-year  
74    project funded by the Seventh Framework Programme (FP7) of the European Union that ran between  
75    November 2013 and October 2016. The overall aim of the project was to establish a sustainable and  
76    economically-feasible strategy for the production of novel high-value phenolic compounds isolated from  
77    berry fruits using bacterial platforms. The project aimed at covering all stages of the discovery and pre-  
78    commercialization process, including berry collection, screening and characterization of their bioactive  
79    components, identification and functional characterization of the corresponding biosynthetic pathways, and  
80    construction of Gram-positive bacterial cell factories producing phenolic compounds. Further activities  
81    included optimization of polyphenol extraction methods from bacterial cultures, scale-up of production by  
82    fermentation up to pilot scale, as well as societal and economic analyses of the processes. This review article  
83    summarizes some of the key findings obtained throughout the duration of the project.

84

## 85    **Introduction**

86    BacHBerry (BACterial Hosts for production of Bioactive phenolics from bERRY fruits, [www.bachberry.eu](http://www.bachberry.eu))  
87    has received funding from the European Commission's Seventh Framework Programme (FP7) under the  
88    theme KBBE.2013.3.1-01: "*Plant High Value Products - from discovery to final product*" with the goal of  
89    bringing together experts from the fields of plant biology, industrial biotechnology, analytical chemistry, and  
90    social sciences in order to assemble and validate a complete pipeline for discovery and microbial production  
91    of novel phenolic compounds of plant origin. The consortium consisted of twelve academic groups, five  
92    small and medium sized enterprises and one large enterprise from eleven different countries, namely Austria,  
93    Chile, China, Denmark, France, Germany, The Netherlands, Portugal, Russia, Switzerland and the United  
94    Kingdom. The diversity of partners has allowed access to a variety of expertise and to plant material that had  
95    not been characterized extensively.

## 96    **Polyphenols as a source of bioactive molecules**

97    Historically, plants have been not only important sources of nutrients and energy in the human diet but also  
98    have provided the fundamentals of traditional and modern medicine. The *Plantae* kingdom has been  
99    estimated to produce more than 200,000 different chemical compounds (Weckwerth 2003), including  
100    compounds with proven and potential medical applications. This is reflected in the high number of  
101    pharmaceutical products based on, or derived from, plant natural products, such as aspirin, morphine,  
102    quinine, paclitaxel and artemisinin (Cragg and Newman 2013). Among the different classes of specialized  
103    (i.e. secondary) metabolites phenylpropanoids stand out due to their diversity and ubiquity in the plant  
104    kingdom. Fleshy fruits, a rich source of bioactive components, such as ascorbate (vitamin C), terpenoids and  
105    polyphenols, are considered to have one the highest antioxidant capacity of commonly eaten foods  
106    (Halvorsen et al. 2006). Polyphenol content and composition can vary substantially between fleshy-fruit  
107    varieties (Carvalho et al. 2013; Miladinović et al. 2014; Kula et al. 2016), different cultivation conditions  
108    (Josuttis et al. 2013; Vagiri et al. 2013; Mazur et al. 2014), harvest times and as a result of different post-  
109    harvest factors (Kårlund et al. 2014; Stavang et al. 2015), but particularly between species (Moyer et al.

2001; McDougall et al. 2008; Ruiz et al. 2010; Marhuenda et al. 2016). Furthermore, plant polyphenols are increasingly being associated with putative bioactivities offering protection against several cardiovascular (Goszcz et al. 2015) and neurological diseases (Figueira et al. *in press*).

The hydroxystilbenes are a group of polyphenols with a C<sub>6</sub>-C<sub>2</sub>-C<sub>6</sub> skeleton with two or more hydroxyl groups, where a central carbon-carbon double bond is conjugated with two phenolic moieties (Quideau et al. 2011; Kasiotis et al. 2013). The most well-known representative of this group is resveratrol (*trans*-3,5,4'-hydroxystilbene) (**Figure 1a**). The compound has gained increasing attention with the discovery of the “French paradox”, an observation that despite a diet with relatively high amounts of saturated fats, French people suffer a relatively low incidence of mortality from coronary heart disease (Renaud and de Lorgeril 1992; Catalgol et al. 2012). Like other polyphenols, stilbenes can be further decorated by *O*-methylation (e.g. resulting in 3-methoxy-resveratrol, or pinostilbene, and 3,5'-dimethoxy-resveratrol, or pterostilbene), acylation (hydrangenic acid) or by glycosylation (e.g. piceid and resveratrolside) (Kim et al. 2002; Becker et al. 2003; Zhang et al. 2009; Wang et al. 2015).

The flavonoids are one of the largest families of phenolic compounds. Flavonoids are characterized by their basic skeleton composed of three rings (Ghosh and Scheepens 2009), including two benzene rings (A & B) and one heterocyclic ring (C) (**Figure 1b**). So far, over 5,000 naturally occurring flavonoids have been characterized from various plants and they have been classified into six subgroups, including anthocyanidins (e.g. cyanidin and delphinidin); flavan-3-ols (e.g. catechin and epicatechin); flavonols (e.g. quercetin and kaempferol); flavones (e.g. luteonin and apigenin); isoflavones (e.g. genistein and daidzein), and flavanones (e.g. hesperetin and naringenin).

Within each subgroup, considerable variation can exist when it comes to phenolic composition of different fruits, and in particular of berries. Anthocyanin composition provides an excellent example of this diversity: the red-orange color of strawberries is due to the presence of pelargonidin-type anthocyanins in the flesh and skin whereas the deep purple-black color of blackcurrants is due to the accumulation of high amounts of delphinidin and cyanidin-type anthocyanins in the skin.

135 Anthocyanidins are flavonoids, which are characterized by a hydroxyl group in position 3 and a C-ring with  
 136 a positively-charged oxonium ion (**Figure 1b**). Anthocyanins are water-soluble glycosides of  
 137 anthocyanidins, in which sugars, consisting of one or more hexoses, are linked to the 3-hydroxyl group of the  
 138 pyrylium ring (Heldt and Piechulla 2011). These compounds are responsible for the orange/red-to-blue  
 139 coloration observed in some plants. The most common anthocyanidins include cyanidin (red), delphinidin  
 140 (bluish-purple), malvinidin (reddish-purple), pelargonidin (orange-red), peonidin (pink) and petunidin  
 141 (purple). The distribution of anthocyanidins can vary greatly among different berry species (**Table 1**).

142 Flavonoid compounds frequently are decorated with sugar moieties, sugar-acyl moieties (Giusti and  
 143 Wrolstad 2003) and can be associated with other flavonoids. The patterns of decoration differ greatly  
 144 amongst species (see Supplementary Tables S1 and S2).

145 Flavonoids are far more prevalent and diverse in berry species than in other common fruits and vegetables.  
 146 High flavonol contents are observed in cranberry, lingonberry and blackcurrant (Häkkinen et al. 1999),  
 147 anthocyanins are the most abundant polyphenol pigments (2-5 g kg<sup>-1</sup> fresh weight) in berries (Määttä et al.  
 148 2001) and many simple phenolic acids are abundant in a wide range of berry species (Herrmann and Nagel  
 149 1989). Additionally, berries constitute one the most important dietary sources of ellagitannins such as  
 150 sanguin H6, lambertianin C (Törrönen 2009; Landete 2011) and condensed tannins such as the  
 151 proanthocyanidins (Rasmussen et al. 2005; Hellström et al. 2009). Stilbenes, in contrast, are not that  
 152 widespread in berries: resveratrol shows highest abundance in grapes (up to 20-30 mg kg<sup>-1</sup> fresh weight), but  
 153 small amounts of resveratrol, pterostilbene and piceatannol have been also detected in blueberry, cowberry,  
 154 cranberry, deerberry and lingonberry (Jeandet et al. 1991; Paredes-López et al. 2010; Manganaris et al.  
 155 2014). It is clear that soft-fruit species contain a staggering diversity of distinct polyphenol derivatives whose  
 156 potential is yet to be harnessed.

157 The market for polyphenolic compounds has seen substantial growth over the past few years, and is expected  
 158 to exceed 850 million USD by 2018 (Aranaz et al. 2016). The major factors contributing to this trend are the  
 159 growing consumer awareness regarding the benefits of polyphenol consumption and the increasing use of



polyphenol-containing extracts in food, beverage, and cosmetics products particularly in Asia (Jain and Ramawat 2013; Aranaz et al. 2016; Grand View Research Inc. 2016). The increase in demand also requires efficient and eco-friendly production processes, to improve on current manufacturing practices that mostly rely on extraction from various plant sources (*e.g.* roots, leaves or fruits) via complex downstream processing (Wang et al. 2016a). The BacHBerry consortium decided to address these challenges and set the following objectives: (i) to systematically analyze the phenolic contents in the large berry germplasm collections available from consortium members, (ii) to establish a publicly-available database of transcriptomic and metabolic data obtained from berry bioprospecting within, as well as outside of the project, (iii) to discover novel bioactivities in berry extracts against a range of human pathologies, such as Alzheimer's disease and Amyotrophic Lateral Sclerosis (ALS), by high-throughput screening with subsequent identification of functional biomolecules, (iv) to identify the corresponding biosynthetic genes and perform functional characterization of the respective gene products, (v) to assay a selection of the biosynthetic genes for functionality in Gram-positive bacterial hosts and use those to construct bacterial cell factories for phenolic production, (vi) to improve the production efficiency further by introducing modifications to the host metabolic networks predicted via rational design or computational tools developed within the project, (vii) to design and optimize cost-effective food-grade methods for extraction of phenolic compounds from bacterial fermentation broth, and (viii) to optimize fermentation conditions and subsequently upscale the production to pilot plant levels.

The project commenced with the selection, sampling and whole-metabolite profiling of berries. In order to explore the potential of the phytochemical diversity present in different berry species, we undertook an untargeted characterization of a wide collection of berries from different species/cultivars utilizing metabolomics-based methods to aid the selection of candidate berry extracts for bioactivity screening.

## **Phytochemical diversity in sampled berries**

Although significant advances have been made in characterizing the polyphenolic complement of berries, particularly in the context of cultivated species and genotypes, there is limited literature available regarding

185 the phytochemical composition of wild and underutilized species/cultivars. Indeed, wild berries are  
186 commonly reported as particularly rich in diverse phenolic compounds, often as a result of phenolic-  
187 associated astringency having been bred out of cultivated species (Häkkinen et al. 1999). The phenolic  
188 diversity of individual berries has been highlighted in many publications (Määttä et al. 2003; Zadernowski et  
189 al. 2005; Milivojević et al. 2011), but studies that capture this broad diversity systematically are limited. In  
190 BacHBerry, we aimed not only to address this knowledge gap but also to go beyond the state-of-the-art by  
191 combining analyses of phenolic diversity with functional testing of berry extracts. The genera targeted in this  
192 study included *Rubus*, *Ribes*, *Vaccinium*, *Lonicera*, *Lycium*, *Aristotelia*, *Berberis* and *Ugni* collected from  
193 different locations in the world including Chile, China, Portugal, Russia and United Kingdom (see  
194 **Supplementary Table S3**).

#### 195 *Berry extract metabolite profiling*

196 Fruit samples from a total of 112 species/genotypes were collected, extracted, and subsequently subjected to  
197 liquid-chromatography mass-spectrometry (LC-MS) metabolomic analysis (see details in Supplementary  
198 Materials S1). An untargeted method was used to generate a total of 1,890 mass spectral features (1,506 and  
199 384 in positive and negative modes, respectively), which were integrated to generate the dataset used for  
200 statistical analysis. Principal component analysis (PCA) was used as a multivariate statistical analysis tool to  
201 identify those species, which had the most distinct phytochemical profiles (**Figures 2a and 2b**).

202 Using the four principal components it was possible to select the species/genotypes, which represented the  
203 broadest phenolic diversity (**Table 2**). In addition to selecting the outlier groups of species, it was also  
204 decided to include a small subset of samples (two *Rubus idaeus* genotypes), which did not separate from the  
205 majority of samples in the first principal components. This provided samples with phenolic profiles distinct  
206 from the outlier samples and standards for comparison between uncultivated/underutilized species and  
207 commonly grown species; in essence a commercial control. The fruits from the selected species (**Table 2**)  
208 were extracted and tested for bioactivity in yeast (*Saccharomyces cerevisiae*) models for pathological  
209 processes associated with several chronic disorders.

## 210 *Berry extract bioactivity screening*

211 Evidence for the protective role of polyphenols against chronic diseases has increased over the past 20 years  
212 (Figueira et al. *in press*; Goszcz et al. 2015). Neurodegenerative diseases (NDs) represent a group of chronic  
213 neurological conditions affecting millions of people worldwide, among which are the Alzheimer's Disease  
214 (AD), the Parkinson's Disease (PD), the Huntington's Disease (HD), and Amyotrophic Lateral Sclerosis  
215 (ALS). All these diseases have in common the aggregation and deposition of protein aggregates, namely of  
216 amyloid  $\beta$ 42 (A $\beta$ 42) (O'Brien and Wong 2011),  $\alpha$ Synuclein ( $\alpha$ Syn) (Shults 2006), huntingtin (HTT)  
217 (Miller-Fleming et al. 2008) and FUS in Sarcoma (FUS) (Kwiatkowski et al. 2009; Vance et al. 2009),  
218 which represent the pathological hallmarks of AD, PD, HD and ALS, respectively. In addition, chronic  
219 activation of innate immune responses is a process closely associated with neurodegeneration. Its modulation  
220 is driven by persistent activation of key transcription factors, such as the Nuclear Factor of Activated T-cells  
221 (NFAT) and Nuclear Factor  $\kappa$ B (NF $\kappa$ B), which upregulate pro-inflammatory responses creating a positive  
222 feedback loop further amplifying initial stimuli. It has been argued that disruption of this loop may represent  
223 an important strategy to mitigate the progression of NDs. The pleiotropic effects of polyphenols have been  
224 shown to interfere with aggregation-driven neurodegeneration as well as to attenuate chronic inflammatory  
225 processes, thereby improving health of cellular and animal models (Figueira et al, *in press*). Consequently,  
226 polyphenol-based therapies represent an underexplored strategy to minimize the huge social and economic  
227 impact of NDs.

228 The high degree of evolutionary conservation of fundamental biological processes among eukaryotes has  
229 established yeast as a validated model organism to decipher the intricacies of human pathologies, particularly  
230 NDs, to identify molecular targets amenable to therapeutic intervention as well as lead molecules with  
231 health-promoting potential (Kritzer et al. 2009; Su et al. 2010; Tardiff et al. 2012; Menezes et al. 2015).  
232 BacHBerry aimed at identifying phenolic bioactives from harnessing the diversity of phenolics in selected  
233 berry germplasm from cultivated, wild and underutilized species of berries. The yeast-based screening  
234 platform for bioactivity identification comprised a set of genetically modified *S. cerevisiae* strains expressing  
235 Green Fluorescent Protein (GFP) fused with A $\beta$ 42 (Bharadwaj et al. 2010),  $\alpha$ Syn (Outeiro and Lindquist

2003), HTT (Krobitsch and Lindquist 2000) and FUS (Ju et al. 2011) under the control of a galactose-inducible promoter (**Figure 3a**). Upon activation of expression, these proteins start forming aggregates, which consequently has a negative impact on cellular growth and results in lower GFP fluorescence levels. In case an extract possesses bioactivity against one of the diseases, addition of the extract to the activated yeast cells reduces growth inhibition in the corresponding model. These yeast strains accelerated the identification of phenolic compounds with health-promoting attributes among the most chemically diverse samples identified by the metabolomic analysis (**Table 2**). For information on the procedures used during the screening see Supplementary Materials S1.

An illustrative example of how bioactivities for AD were identified is given in **Figure 4a**. Upon shift of cells to galactose medium to induce expression of GFP-A $\beta$ 42, yeast growth was impaired in comparison to the control strain indicating GFP-A $\beta$ 42 proteotoxicity. The treatment with polyphenol-enriched extracts of *Lycium chinense* significantly recovered growth of these cells revealing its protective role via modulation of A $\beta$ 42 toxicity. The bioactivities for PD, HD and ALS were screened using a similar approach, in cells expressing the respective disease proteins.

In addition, we have also used a yeast-based model for inflammatory signaling that is based on a Crz1 reporter-strain (Prescott et al. 2012; Garcia et al. 2016) (**Figure 3b**). Crz1 is a yeast orthologue of NFAT, which is an important modulator of inflammation in humans that is known to be involved in development of multiple disorders, such as the inflammatory bowel disease or the rheumatoid arthritis (Pan et al. 2013). Both Crz1 and NFAT are known to be activated by a serine/threonine protein phosphatase calcineurin (CaN) in a Ca<sup>2+</sup>-dependent manner (Rusnak and Mertz 2000; Bodvard et al. 2013). The utilized reporter strain encodes the  $\beta$ -galactosidase gene (*lacZ*) under the control of a promoter bearing Crz1-binding sites, the Calcineurin-Dependent Responsive Element (CDRE), allowing the assessment of Crz1 activation through the measurement of  $\beta$ -galactosidase activity using chromogenic substrates (Garcia et al. 2016). Given the evolutionary conservation of NFAT and Crz1 activation mechanisms, in combination with the conserved activity of the immunosuppressant FK506 both in yeast and in humans, the yeast Ca<sup>2+</sup>/CaN/Crz1 reporter assay represents a reliable tool to identify bioactives with potential to attenuate NFAT-mediated

262 inflammatory responses. The potential of *Lycium chinense* polyphenol-enriched extracts to attenuate  
263 inflammation is shown in **Figure 4b**, exemplifying the approach used in BacHBerry to filter for potential  
264 anti-inflammatory activities. The activation of Crz1 by MnCl<sub>2</sub>, mimicking NFAT activation, led to high β-  
265 galactosidase activity and cell treatment with FK506 and *Lycium chinense* extract strongly reduced the β-  
266 galactosidase levels revealing its ability to modulate Crz1, and potentially NFAT, activation.

267 These examples illustrate the strategy used in the BacHBerry project to search for bioactive compounds  
268 interfering with pathological processes of NDs and inflammation. The yeast-based screening platform also  
269 included strains allowing the identification of metabolites with potential application for type II diabetes,  
270 hematological diseases and cancer (unpublished results).

#### 271 *Bioassay-guided fractionation compound discovery and candidate compound validation*

272 This method of discovering new bioactive natural products ultimately depends on the availability of  
273 biological material and preparative- or semi-preparative-scale analytical methods with the capability of  
274 resolving mixtures of different classes of compounds typically present in berry extracts (Pauli et al. 2012).  
275 Semi-preparative chromatography was used to fractionate selected berry extracts with potential bioactive  
276 properties using a hybrid approach of bioassay-guided fractionation procedure (Yang et al. 2001; Weller and  
277 G. 2012; Tayarani-Najaran et al. 2013) and screening of pure compounds (Watts et al. 2010). Bioassay-  
278 guided fractionation typically involves the following steps: assessment of bioactivity, extraction of the  
279 biological material with different solvents and testing of bioactivity. Once bioactivity of an extract has been  
280 validated, the extract gets subjected to an iterative process of sub-fractionation/bioactivity testing until pure  
281 bioactive natural products are obtained for structural characterization. This approach benefits from exclusion  
282 of extracts that do not have bioactivity. However, this procedure requires extensive use of biological material  
283 and expensive materials and may result in the isolation of already-known natural products (Duarte et al.  
284 2012). Another disadvantage is that the approach is based on the assumption that bioactivity is conferred by  
285 a pure compound, although this method can also be used for identification of bioactivities conferred by a  
286 cumulative interactions of several polyphenols. Alternatively, pure-compound screening relies on an initial

isolation and structural elucidation of the individual compounds present in the biological extract followed by bioassay screening. This strategy allows the researcher to focus solely on novel compounds, without re-discovering compounds with well-annotated bioactivity. However, it may also lead to the identification of novel compounds with no bioactivity (Duarte et al. 2012). The second method misses any synergistic interactions affecting bioactivities of berry phenolics.

The limited amount of biological material available restricted the number of iterations of fractionation and re-testing of fraction bioactivity possible for typical bioassay-guided fractionation approaches. As described previously, berry extracts typically comprise a relatively diverse and large pool of metabolites that surpass by far the throughput capability of the bioactivity assays used in this study. To overcome these challenges, a hybrid approach was adopted which consisted of several steps: (i) assessment of potential bioactivity present in extracts (as described above), (ii) fractionation of bioactive extracts (see Supplementary Materials S1), (iii) assessment of potential bioactivity present in fractions, (iv) mass-spectrometry-based chemical characterization of bioactive fractions, (v) bioactivity testing of pure candidate compounds (**Figure 5**). Although this approach shared some of the limitations of the other approaches, it did allow for the exclusion of non-bioactive biological extracts or fractions and focused on the identification novel compounds with potential bioactivities with limited requirement for biological material.

Following fractionation of the most promising berry extracts from the first round of screening using the yeast-based platform, isolated fractions were re-analyzed in order to detect bioactive fractions. Subsequently, metabolomic analyses were used for the identification of the individual compounds in each of the bioactive fractions. By utilizing this approach, a set of candidate compounds was defined and re-tested as pure compounds in the yeast-based platform, with further bioactivity validation in the mammalian cell models. Following the BacHBerry pipeline, this information then served as starting point for the engineering of the synthetic pathway for biosynthesis of one of the validated compounds in food-grade bacteria and its subsequent production in small-scale fermenters. The final step of the pipeline, namely re-confirmation of bioactivity of microbially-produced compounds extracted from fermentation broth in both yeast and mammalian cell models, is currently underway.

### 313 Identification of metabolic pathways and regulators involved in phenolic production

314 The next step within the BacHBerry pipeline involved the identification of the biosynthetic genes and  
315 pathways corresponding to the newly discovered bioactives, in order to establish production of these  
316 compounds in bacterial hosts. In plants, the biosynthesis of polyphenols occurs via the phenylpropanoid  
317 pathway, where the aromatic amino acid L-phenylalanine serves as a precursor. Depending on the enzyme  
318 combination, several compounds can be produced in bacteria, including cinnamic acid derivatives, lignin  
319 subunits, lignans, phenylpropenes and coumarins, all sharing a basic C<sub>6</sub>-C<sub>3</sub> skeleton (Vogt 2010).  
320 Furthermore, the phenylpropanoid backbone can be extended with up to three acetyl-units derived from  
321 malonyl-CoA by chalcone synthase (CHS) or stilbene synthase (STS), polyketide synthases which generate  
322 various polyketides, such as flavonoids and stilbenes. The core pathways of phenolic biosynthesis are very  
323 well understood, with common enzymes, such as, phenylalanine ammonia lyase (PAL), cinnamate 4-  
324 hydroxylase (C4H), *p*-coumaroyl CoA ligase (4CL), hydroxycinnamoyl-CoA shikimate/quinate  
325 hydroxycinnamoyltransferase (HCT), *p*-coumaroyl shikimate/quinate 3'-hydroxylase (C3'H), stilbene  
326 synthase and chalcone synthase, found in many plant species (**Figure 6**).

327 Phenolic chemodiversity in berries, as in other plants, derives from a variety of decorations to the backbone  
328 of phenolic structures that include hydroxylations, *O*-methylations, prenylations, aryl migrations,  
329 glycosylations, acylations and polymerizations. For example, anthocyanins are glycosylated on 3-*OH*  
330 positions of the C-ring (**Figure 1**) and many harbor additional glycosyl groups (for example 5'-*O*-  
331 glycosylation and 3'-*O*-glycosylation), methoxy groups (petunidin and malvinidin), or hydroxyl groups  
332 (which differentiate pelargonidin, cyanidin and delphinidin (**Figure 6 and Table 1**)). These decorations may  
333 affect the bioavailability, bioactivity and stability of polyphenolic compounds. It is therefore essential to  
334 develop a greater knowledge of the decorating enzymes specific to berry bioactives, such as those  
335 differentially decorating anthocyanins, flavonols, stilbenes, or condensed tannins. Many genes encoding  
336 enzymes involved in flavonol and anthocyanin decoration have already been characterized not only *in*  
337 *planta*, but also heterologously in microorganisms, such as *Escherichia coli* and *S. cerevisiae* (Tohge et al.  
338 2005; Luo et al. 2007; Pandey et al. 2016; Wang et al. 2016a).

339 Berry bioactives identified within the project had special decorations, and those where the genes encoding  
340 the decorating activities had not been previously identified, required further research. In a recent work, the  
341 transcriptome of the commercial blueberry (*Vaccinium corymbosum*) was studied using RNA sequencing  
342 technology (RNA-seq) in order to discover putative genes related to antioxidant production using next-  
343 generation sequencing (NGS) (Li et al. 2012). Similarly, we aimed to identify decorating enzymes and  
344 regulators of the biosynthetic pathways of berry bioactives by obtaining transcriptome sequences of target  
345 species at different stages of fruit development. This information was used to construct a database of  
346 transcriptome sequences for a broad range of berry species, especially those identified as target species  
347 containing the promising bioactive molecules. In addition, the overall production of polyphenols is regulated  
348 by a range of transcription factors and these were also investigated within the BacHBerry project, with the  
349 objective of identifying markers for breeding enhanced levels of phenolic bioactives using the diverse  
350 germplasm in selected berry species.

351 Firstly, a library of reference sequences consisting of plant gene sequences encoding enzymes of the core  
352 phenolic biosynthetic pathways, which was subsequently used for BLAST searches of berry transcriptomes,  
353 was compiled. This library included well-known and experimentally validated anthocyanin, flavonol,  
354 stilbene and catechin biosynthetic enzymes from a wide range of plant species. Next, 13 germplasms  
355 spanning eight genera, seven families and seven orders were selected for RNA-seq of berry transcriptomes.  
356 Sequencing of selected berries at different stages of maturation was also conducted. The data have provided  
357 significant resources not only for understanding phenolic biosynthesis in berries, but also for investigating  
358 commonalities in genes expressed in berries (**Table 3**).

359 Multi-layer comparisons of the metabolomics and transcriptomics data of developmental stages, and  
360 targeted, as well as, non-targeted searches of candidate transcripts were performed to elucidate genes  
361 involved in specific phenolics biosynthetic pathways. Analyses of expression of these genes from different  
362 developmental stages of selected berry species were performed with the expVIP tool (Borrill et al. 2016).



363 These data were used to search for genes encoding putative decorating enzymes based on the list of  
364 molecules of potential interest as bioactives and having special decorations in target berry species.  
365 Candidates were identified mostly by searching for homologues of well-characterized plant genes. This was  
366 carried out with the help of MassBlast, a software developed within the BacHBerry project  
367 (<https://github.com/averissimo/mass-blast>; (Veríssimo et al. *submitted*)). We annotated all major enzyme  
368 families, including cytochromes P450 (CYP), 2-oxoglutarate-dependent dioxygenases (2-OGDs), UDP-  
369 glycosyltransferases (UGTs), BAHD acyltransferases, and *O*-methyltransferases (OMTs), in all 13 selected  
370 berry germplasms. Average counts per transcriptome are presented in **Table 4**. We also detected candidate  
371 transcript(s) for 60-80% of the known biochemical steps of the phenylpropanoid pathway, representing  
372 between 109-180 candidates per transcriptome. Approximately 40% of these candidates identified from the  
373 transcriptomes were full-length sequences. Furthermore, biochemical activities of some of enzymes encoded  
374 by the candidate genes were tested.

375 We also looked for transcripts encoding regulators of phenolic biosynthesis. Since the R2R3-MYB gene  
376 family operates extensively in controlling phenylpropanoid metabolism in plants (Stracke et al. 2001;  
377 Hartmann et al. 2005), initial searches were focused on identifying MYB-genes that were strongly expressed  
378 in berries. However, members of the bHLH and bZIP families of transcription factors were also identified in  
379 the berry RNA-seq databases since they may also play roles in regulating accumulation of phenolic  
380 bioactives (Hartmann et al. 2005). These data could be particularly useful in breeding programs for selection  
381 of plants with improved bioactive contents. The presence of genes encoding putative transporters of phenolic  
382 compounds was also examined. Excretion of target compounds produced by bacterial hosts into the culture  
383 medium via such transporters would simplify extraction procedures, and may also be necessary in case of  
384 any toxic effects of these molecules in bacteria. One transporter is currently being characterized functionally.

### 385 **Gram-positive bacteria as cell factories for the efficient production of berry high-value phenolics**

386 While the search for novel bioactives was under way, a set of well-characterized biosynthetic pathway was  
387 used for establishing a platform for production of phenolic compounds in Gram-positive bacteria. Two major

388 criteria were used for the selection of host organisms for this task: (1) they had to be robust, well-studied  
389 organisms with a long-standing history of industrial use and (2) the bacteria must have a generally regarded  
390 as safe (GRAS) status and/or be recognized as food-grade. Two industrially-important bacteria, namely  
391 *Lactococcus lactis* and *Corynebacterium glutamicum*, appeared as attractive candidates for the sustainable  
392 production of added-value food and pharmaceutical compounds (Lahtinen et al. 2011; Jojima et al. 2013).  
393 Both *L. lactis* and *C. glutamicum* are used worldwide in numerous industrial processes, which consequently  
394 led to significant amount of research work dedicated towards improving our knowledge of their physiology  
395 and genetics. These efforts have resulted in development of large toolboxes for (heterologous) gene  
396 expression and introduction of genetic modifications in these bacteria (Burkovski 2008; Pontes et al. 2011;  
397 Gaspar et al. 2013). Furthermore, high-throughput methodologies for transcriptome, proteome and  
398 metabolome analysis, and genome-scale metabolic models are available for these species (Krömer et al.  
399 2004; Oliveira et al. 2005; Shinfuku et al. 2009; Gaspar et al. 2013). Combined with their general robustness,  
400 ease of handling and the GRAS status, both *L. lactis* and *C. glutamicum* have a strong potential for becoming  
401 platform organisms for production of polyphenolic compounds.

402 It should also be noted that during the earlier stages of the project, *E. coli* was also as a platform for  
403 reconstruction of polyphenol biosynthetic pathways. *E. coli* has a long history of use in research on  
404 microbial-based production of various value-added chemicals. Moreover, it is one of the best studied  
405 organisms, easy to handle, and there much larger set of genetic tools is available for this organism compared  
406 to other bacteria (Yu et al. 2011; Dobson et al. 2012; Chen et al. 2013). Furthermore, *E. coli* has been  
407 previously used for heterologous production of polyphenolic compounds (Pandey et al. 2016; Wang et al.  
408 2016a).

#### 409 *Lactococcus lactis*

410 *L. lactis* is a Gram-positive bacterium from the order of *Lactobacillales* that has been widely used in the food  
411 industry for the manufacturing of cheese, buttermilk and other dairy products. It is a low G+C aerotolerant  
412 bacterium with a relatively small genome and simple sugar metabolism: utilization occurs via glycolysis  
413 (Embden–Meyerhof–Parnas pathway) with lactic acid being the major end-product (Gaspar et al. 2013). Its

414 ability to ferment various carbohydrates, such as hexoses, pentoses and, in some strains, even complex  
 415 carbohydrates of plant origin, as well as its high alcohol, acid and stress tolerance makes this bacterium  
 416 suitable for the harsh environments commonly found under industrial production conditions. *L. lactis* strains  
 417 have been engineered to produce multiple commodity chemicals (e.g. 2,3-butanediol and butanol), as well as  
 418 food additives, including sweeteners (e.g. mannitol and xylitol), flavorings (e.g. acetaldehyde and diacetyl),  
 419 and vitamins. Furthermore, production of plant secondary metabolites, namely isoprenoids and stilbenes, has  
 420 been demonstrated in this organism (Neves et al. 2005; Katz et al. 2011; Gaspar et al. 2013; Song et al.  
 421 2014). In addition, *L. lactis* is well known as an excellent host for the expression of both prokaryotic and  
 422 eukaryotic proteins, and has been consequently used for development of live vaccine (Kunji et al. 2005;  
 423 Hernández et al. 2007; Frelet-Barrand et al. 2010; Pontes et al. 2011).

424 As depicted in **Figure 6**, in plants the biosynthesis of the core intermediate of the phenylpropanoid pathway,  
 425 *p*-coumaric acid, begins from *L*-phenylalanine and occurs in two steps: a deamination reaction catalyzed by  
 426 phenylalanine ammonia-lyase (PAL) and a hydroxylation reaction catalyzed by cinnamate 4-hydroxylase  
 427 (C4H). The latter enzyme belongs to the CYP family that is often challenging to express in active form in  
 428 prokaryotic systems (Dvora and Koffas 2013). However, in some bacteria and fungi there exists an  
 429 alternative pathway where *L*-tyrosine is converted directly into *p*-coumaric acid via the action of tyrosine  
 430 ammonia-lyase (TAL), hence bypassing the C4H step (**Figure 6**). As a first step towards establishing a  
 431 platform strain for the production of polyphenols in *L. lactis*, in collaboration with Nielsen and coworkers,  
 432 we identified several novel TAL enzymes that specifically produce high levels of *p*-coumaric acid from *L*-  
 433 tyrosine (Jendresen et al. 2015).

434 To test the potential of *L. lactis* to produce more complex phenolic compounds, we introduced a pathway for  
 435 the biosynthesis of resveratrol. The introduced pathway consisted of three enzymes: TAL, 4CL (4-  
 436 coumaroyl-CoA ligase, which converts *p*-coumaric acid to *p*-coumaroyl-CoA), and STS (stilbene synthase,  
 437 which catalyzes the condensation of *p*-coumaroyl-CoA and three molecules of malonyl-CoA to yield *trans*-  
 438 resveratrol, **Figure 6**). The corresponding genes were placed on the high copy number vector pNZ8048  
 439 under the control of the nisin-inducible promoter  $P_{nisA}$ . This promoter is a part of a tightly-controlled

inducible gene expression system which is based on the auto-regulatory mechanism controlling the production of the bacteriocin nisin (Kuipers et al. 1993; Kuipers et al. 1998). The producer *L. lactis* strain contains a chromosomally-integrated two-component system NisRK allowing induction of its target promoters by addition of sub-inhibitory concentrations of the bacteriocin nisin to the exponentially-growing cell culture (Linares et al. 2010). Upon cultivation of the producer strain in a chemically defined medium, production of *trans*-resveratrol was detected in the order of several mg l<sup>-1</sup>. When TAL was replaced with a promiscuous PAL/TAL enzyme (MacDonald and D’Cunha 2007), additional production of small amounts of pinosylvin (*trans*-3,5-dihydroxystilbene) from L-phenylalanine was observed. Furthermore, production of methylated variants of *trans*-resveratrol was also achieved (unpublished results). Obtained titers for *trans*-resveratrol are comparable to that obtained in an early work on *trans*-resveratrol production in both *E. coli* and yeast (Beekwilder et al. 2006; Choi et al. 2011). Current research efforts focus on improving the production efficiency through metabolic engineering of the central carbon metabolism for increased precursor supply.

Anthocyanins (glycosylated anthocyanidins) are another group of phenolic phytochemicals. In addition to their antioxidant properties, anthocyanins are valuable for their colors, which range from orange to blue (Falcone Ferreyra et al. 2012). They are of particular interest for the food industry as natural alternatives to replace synthetic dyes. Anthocyanins are synthesized from flavanones by flavanone 3-hydroxylase (F3H), which adds a hydroxyl group to the C3 position. Resulting dihydroflavonols are further processed by dihydroflavonol 4-reductase (DFR), which catalyzes reduction of the carbonyl group of the C4 position to form leucoanthocyanidins, which in turn are converted to anthocyanidins via the action of anthocyanidin synthase (ANS). Anthocyanidins are further stabilized through a C3-glycosylation reaction usually catalyzed by UDP-glucose:flavonoid 3-*O*-glucosyltransferase (UFGT) to form the anthocyanidin 3-*O*-glucosides (Falcone Ferreyra et al. 2012) (**Figure 6**).

In order to reconstruct the anthocyanin production pathway in *L. lactis*, native genes and genes codon-optimized for *L. lactis* from various plant species coding for F3H, DFR, ANS and UFGT were placed under control of the nisin-inducible promoter P<sub>nisA</sub> on a high copy-number vector or were integrated into the

466 chromosome of *L. lactis*. Each protein was tagged with an N- or C- terminal strep-tag, and Western blots  
467 were performed to detect the heterologous proteins in cell-free extract of *L. lactis*. Activity of each enzyme  
468 was probed *in vitro* and *in vivo* and the reaction products were analyzed using high performance liquid  
469 chromatography (HPLC). Production of various types of cyanidin-derived anthocyanidins and anthocyanins  
470 by the engineered *L. lactis* strains was detected, further demonstrating the potential of this lactic acid  
471 bacterium as a production host for plant-derived bioactive compounds for food applications (unpublished  
472 results). To our knowledge, this is the first report of using lactic acid bacteria for the production of complex  
473 phenolic compounds.

474 Malonyl-CoA is a rate-limiting precursor in the biosynthesis of various phenolic compounds in bacteria (Xu  
475 et al. 2011; Bhan et al. 2013). Under physiological conditions, *L. lactis* cells use malonyl-CoA as a substrate  
476 in the chain elongation step of the fatty acid biosynthesis. FapR of *Bacillus subtilis* is a transcriptional  
477 repressor that directly senses the intracellular pool of malonyl-CoA. A metabolic biosensor based on the  
478 interaction of malonyl-CoA with FapR was previously designed in *Escherichia coli* (Xu et al. 2013). As a  
479 part of the efforts to enhance polyphenol production, a transcriptional biosensor for malonyl-CoA was  
480 constructed for the use in *C. glutamicum* but also proved to be fully functional in *L. lactis* and could  
481 therefore be used for screening of variants with an enhanced malonyl-CoA pool.

482 Among other activities that were pursued in the early stages of the project, we also tried to express plant  
483 CYPs in *L. lactis*, as many steps of the polyphenol biosynthetic pathway are catalyzed by these heme-bound  
484 monooxygenases. Functional expression of genes encoding membrane-anchored CYPs in prokaryotes is  
485 known to be notoriously difficult, and is limited by problems associated with the differences in intracellular  
486 localization (natively to the plant endoplasmatic reticulum), membrane composition, protein folding, and  
487 cofactor utilization. The latter element became especially crucial when attempting to express these proteins  
488 in *L. lactis*, because this, normally anaerobic, bacterium does not have the ability to synthesize heme, despite  
489 undergoing a metabolic shift to respiration when supplemented with an exogenous source of heme and being  
490 capable of regulating heme homeostasis (Pedersen et al. 2012).

491 We focused our efforts on CYP79A1 from *Sorghum bicolor* as it is a well-characterized CYP involved in  
492 dhurrin biosynthesis (Bak et al. 2000), and has been previously successfully produced in *E. coli* (Vazquez-  
493 Albacete et al. 2016). As a testbed for the expression of this CYP we have used the Waldo platform (Waldo  
494 et al. 1999) that was previously instrumental in producing functional CYP79A1 in *E. coli*. Briefly, the  
495 platform is based on a fusion of the membrane protein with a GFP folding reporter at the C terminus: the  
496 fused GFP is designed to act as a fluorescent reporter only when the membrane protein is properly folded  
497 and its C terminus is oriented towards the cytoplasm.

498 We detected CYP79A1-GFP by both fluorescence measurements of the *L. lactis* culture and in-fluorescence  
499 protein gels, but were not able to detect activity of the CYP in our *in vitro* tests. Purified membrane protein  
500 fractions from cell cultures grown both aerobically (with hemin supplementation) and anaerobically (with  
501 addition of hemin after purification of membrane proteins) did not show any product formation in our TLC-  
502 based assay (Blomstedt et al. 2012). These results suggested that CYP79A1 was probably misfolded in *L.*  
503 *lactis*, although we did not perform any follow-up experiments to confirm this hypothesis. The topic was not  
504 pursued any further because any CYP active in the biosynthetic pathways for the selected model compounds  
505 could be by-passed (for example the C4H-catalyzed step could be by-passed by using TAL instead of PAL).

#### 506 *Corynebacterium glutamicum*

507 *C. glutamicum* is an important organism in industrial biotechnology for the microbial production of bulk  
508 chemicals, in particular of amino acids (Becker and Wittmann 2012). *C. glutamicum* is a very robust  
509 microorganism, which shows high resistance to the presence of small aromatic compounds (Liu et al. 2013),  
510 which renders *C. glutamicum* a very promising host for the production of pharmacologically interesting  
511 plant-derived polyphenols (Marienhagen and Bott 2013). Until recently, the activity of a complex catabolic  
512 network for aromatic compounds meant that *C. glutamicum* was not used for the production of aromatic  
513 compounds (except for aromatic amino acids) (Shen et al. 2012) due to rapid degradation of the products of  
514 interest. However, *C. glutamicum* can be easily modified genetically and a toolbox for the high-level  
515 expression of heterologous genes originating from other organisms, has become available, and is based on a

516 set of strong and inducible promoters (Eggeling and Bott 2005; Pátek and Nešvera 2013; Kortmann et al.  
 517 2015).

518 Surprisingly, in initial experiments, it turned out that *C. glutamicum* is able to grow on naturally occurring  
 519 phenylpropanoids such as *p*-coumaric acid, ferulic acid, caffeic acid and 3-(4-hydroxyphenyl)-propionic acid  
 520 as sole carbon and energy sources. The maximum growth rates on the selected phenylpropanoids ranged  
 521 from 0.15 to 0.23 h<sup>-1</sup> corresponding to doubling times of 3 to 4.5 hours (Kallscheuer et al. 2016a). We  
 522 suspected that the activity of this pathway would come into conflict with any attempt to produce polyphenols  
 523 as both pathways compete for phenylpropanoids. Unfortunately, the underlying metabolic pathway  
 524 responsible for the degradation of phenylpropanoids in *C. glutamicum* was unknown at the start of the  
 525 BacHBerry project. Global gene expression analyses were conducted, which led to the identification of a  
 526 gene cluster consisting of seven genes (designated *phdT*, *phdA*, *phdR*, *phdB*, *phdC*, *phdD* and *phdE*). In  
 527 response to feeding of the above-mentioned phenylpropanoids these genes were upregulated up to 100-fold  
 528 (Kallscheuer et al. 2016a). In subsequent studies, it turned out that the genes in this cluster code for the  
 529 enzymes of the phenylpropanoid catabolic pathway in *C. glutamicum*. The gene *phdT* codes for a  
 530 phenylpropanoid transporter protein, whereas *phdA* and the predicted operon *phdBCDE* code for enzymes  
 531 involved in the degradation of phenylpropanoids. The *phd* gene cluster is transcriptionally controlled by a  
 532 MarR-type repressor encoded by *phdR*. Cultivation experiments conducted with *C. glutamicum* strains  
 533 carrying single-gene deletions showed that loss of *phdA*, *phdB*, *phdC* or *phdE* abolished growth of *C.*  
 534 *glutamicum* with all tested phenylpropanoid substrates. The intracellular accumulation of pathway  
 535 intermediates determined via LC-ESI-MS/MS in single-gene deletion mutants showed that the *phd* gene  
 536 cluster encodes for enzymes involved in a CoA-dependent,  $\beta$ -oxidative deacetylation pathway, which is  
 537 essential for the utilization of phenylpropanoids in *C. glutamicum*. The identified pathway thereby represents  
 538 a peripheral catabolic route responsible for chain shortening of phenylpropanoids to benzoic acids and  
 539 acetyl-CoA (**Figure 7**). *C. glutamicum* converts the resulting intermediates benzoate, 4-hydroxybenzoate and  
 540 protocatechuate to succinyl-CoA and acetyl-CoA by central degradation pathways for aromatic compounds  
 541 (especially the well-characterized  $\beta$ -ketoadipate pathway), and catabolites are ultimately channeled into the

542 tricarboxylic acid cycle (Shen and Liu 2005) (**Figure 7**). Unexpectedly, *C. glutamicum* failed to grow on the  
543 phenylpropanoid, cinnamic acid, which lacks ring *para*-hydroxylation. Astonishingly, a *C. glutamicum* strain  
544 with a single deletion of the gene *phdR* coding for the identified transcriptional repressor was able to grow  
545 with cinnamic acid as sole carbon and energy source (Kallscheuer et al. 2016a). This indicated that PhdR has  
546 a pronounced specificity for hydroxylated phenylpropanoids, especially when hydroxylated in *para*-position  
547 of the aromatic ring. In the course of our studies, cinnamic acid was the only phenylpropanoid tested lacking  
548 this hydroxylation, so the complete specificity of this catabolic pathway of phenolics in *C. glutamicum*  
549 remains unknown.

550 During construction of the *C. glutamicum* platform strain for plant polyphenol production, four gene clusters,  
551 comprising 21 genes, including four genes of the *phd* cluster involved in the catabolism of aromatic  
552 compounds, were deleted. The resulting platform strain *C. glutamicum* DelAro<sup>4</sup> was unable to degrade any  
553 phenylpropanoid tested, and also could not utilize simple benzoic acids such as protocatechuate (Kallscheuer  
554 et al. 2016c). This strain was used as a chassis for the microbial production of polyphenols in particular,  
555 stilbenes and (2*S*)-flavanones, in *C. glutamicum*.

556 Plasmid-borne expression of codon-optimized genes coding for a 4CL from parsley (*Petroselinum crispum*),  
557 a chalcone synthase (CHS) and a chalcone isomerase (CHI) from *Petunia x hybrida* from the strong T7  
558 promoter in this strain background enabled production of (2*S*)-flavanones (**Figure 7**). Maximal titers of 35  
559 mg l<sup>-1</sup> naringenin and 37 mg l<sup>-1</sup> eriodictyol were obtained from the supplemented phenylpropanoids *p*-  
560 coumaric acid and caffeic acid, respectively. Expression of the *4cl* gene in combination with a codon-  
561 optimized gene coding for an STS from peanut (*Arachis hypogaea*) allowed for the production of the  
562 stilbenes pinosylvin, *trans*-resveratrol and piceatannol starting from supplemented phenylpropanoids  
563 cinnamic acid, *p*-coumaric acid and caffeic acid, respectively (Kallscheuer et al. 2016c) (**Figure 7**). Stilbene  
564 concentrations of up to 158 mg l<sup>-1</sup> could be achieved in defined CGXII medium with 4% glucose and  
565 supplemented with cerulenin, a fatty-acid biosynthesis inhibitor that is known to enhance polyphenol  
566 production in bacteria by increasing the intracellular concentrations of malonyl-CoA (Santos et al. 2011; van  
567 Summeren-Wesenhagen and Marienhagen 2015; Cao et al. 2016). Engineering amino acid metabolism for an



568 optimal connection to the synthetic plant polyphenol pathways was conducted to enable stilbene production  
569 directly from glucose. Indeed, the additional heterologous expression of *aroH* from *E. coli* (coding for the  
570 first enzyme of the shikimate pathway responsible for the synthesis of aromatic amino acids) together with a  
571 codon-optimized gene coding for a TAL from *Flavobacterium johnsoniae* led to the production of up to 60  
572 mg l<sup>-1</sup> *trans*-resveratrol in the absence of supplemented *p*-coumaric acid (Kallscheuer et al. 2016c).  
573 Furthermore, *C. glutamicum* was used to produce *trans*-resveratrol from a cheaper precursor 4-  
574 hydroxybenzoic acid (HBA) through reversal of a microbial  $\beta$ -oxidative phenylpropanoid degradation  
575 pathway (Kallscheuer et al. 2016b). This novel synthetic pathway circumvents any need for the endogenous  
576 supply of aromatic amino acids as polyphenol precursors and in *C. glutamicum* instantly yielded 5 mg l<sup>-1</sup>  
577 *trans*-resveratrol from supplemented HBA without any further optimization. Very recently, we also  
578 engineered *C. glutamicum* to produce more complex polyphenols including flavonols (kaempferol and  
579 quercetin) and *O*-methylated stilbenes (pinostilbene and pterostilbene). The obtained pterostilbene titer was  
580 42 mg l<sup>-1</sup>, which is comparable to the titer of 50 mg l<sup>-1</sup> obtained in engineered *E. coli*. In *C. glutamicum*, for  
581 the flavonols kaempferol and quercetin titers of 23 mg l<sup>-1</sup> and 10 mg l<sup>-1</sup>, respectively, were detected. The  
582 aforementioned titers were obtained from the phenylpropanoids *p*-coumaric acid and caffeic acid ,  
583 respectively, as precursors and represent the highest flavonol titers obtained in engineered microorganism  
584 until today (Kallscheuer et al. 2017).

585 The main bottlenecks for the production of polyphenols turned out to be the TAL activity as well as the  
586 intracellular availability of malonyl-CoA, representing the cosubstrate in the STS- and CHS-reactions. On-  
587 going efforts to improve polyphenol production in *C. glutamicum* therefore focus mainly on improving the  
588 activity of the rate-limiting enzymes and increasing the intracellular malonyl-CoA pool.

#### 589 *Other activities*

590 As a part of this work, we also aimed at reconstruction of a previously unknown biosynthetic pathway for the  
591 flavonol fisetin (3,7,3',4'-tetrahydroxyflavone, **Figure 8**). Fisetin is a potent anti-inflammatory, anti-cancer,  
592 and anti-oxidant compound (Suh et al. 2008; Funakoshi-Tago et al. 2011; Jash and Mondal 2014). Moreover,

593 it has been demonstrated to have a neuroprotective effect and also to limit the complications of type I  
594 diabetes (Maher 2009; Maher et al. 2011). Based on the molecular structure of this compound, we  
595 hypothesized that its biosynthesis should follow the general pathway of flavonol biosynthesis (**Figures 6 and**  
596 **8**): a flavanone precursor would be converted into dihydroflavonol by F3H, and then further converted into  
597 the corresponding flavonol via the action of flavonol synthase (FLS). In the last step of the biosynthesis, a  
598 hydroxyl group would be added at the C3' position by flavonoid 3'-hydroxylase (F3'H). This part of the  
599 work was carried out in *E. coli*. Isoliquiritigenin (2',4',4-trihydroxychalcone) appeared to be the most suitable  
600 precursor, and when supplied to an *E. coli* culture expressing F3H and FLS from *Arabidopsis thaliana*,  
601 production of garbanzol and resokaempferol was detected. Resokaempferol could be further converted into  
602 fisetin by expression of F3'H genes from *A. thaliana* and *Pn. hybrida* fused with a cytochrome-P450  
603 reductase (CPR) from *Catharanthus roseus* (Stahlhut et al. 2015). The three enzymes were then combined  
604 with the isoliquiritigenin biosynthetic pathway, consisting of TAL from *Rhodobacter sphaeroides*, 4CL from  
605 *Pr. crispum*, CHI from *Medicago sativa*, and a fusion protein of CHS from *Pn. hybrida* and CHR from  
606 *Astragalus mongholicus*, to yield 0.3 mg l<sup>-1</sup> of fisetin (Stahlhut et al. 2015). This work allowed reconstruction  
607 of a complete pathway for the biosynthesis of fisetin from L-tyrosine as the starting precursor using enzymes  
608 from various sources, and the identified pathway is currently being transferred to Gram-positive hosts.

## 609 **Bioinformatics and modelling**

610 There were multiple activities within BacHBerry, both within the discovery and the microorganism  
611 engineering sectors, which required the use of bioinformatics tools. Hence, the mathematical modeling and  
612 computational parts of the project aimed at addressing three major problems. The first two were related to  
613 gene expression, either for performing functional genomics analysis using the RNA-seq data from berries or  
614 to design sequences for heterologous expression of plant genes in microbes. The third problem concerned  
615 metabolic network optimization for both a single population and for a community of microbes. We detail  
616 below the main results obtained on each of the above activities. Other subjects that were addressed as a part  
617 of this project were metabolic communication between the different species in a community (Milreu et al.  
618 2014; Andrade et al. 2016) and modeling dynamic metabolic networks for potential application in strain

619 optimization (Costa et al. 2014; Hartmann et al. 2015a; Hartmann et al. 2015b; Costa et al. 2016; Hartmann  
620 et al. 2016).

### 621 *Sequencing data analysis and plant functional genomics*

622 Understanding which genes are differentially expressed during phenolic production by plants requires the  
623 capacity to undertake a rapid analysis of RNA-seq data. One of the main problems to be addressed in this  
624 case is the search for coding sequences that are homologous to those already present in sequence databases,  
625 which involves identification, selection and annotation of the best matches. Currently, such analyses rely  
626 heavily on manual curation as well as on the manual assembly of results obtained using different  
627 computational procedures and sequence databases. Automated processes become crucial in the light of the  
628 ever-increasing amounts of generated transcriptome data. Furthermore, automation of processes also allows  
629 improvement of the quality of the analyses, as the likelihood of human error grows with the amount of  
630 manual processing and verifications necessary. Moreover, most of the existing tools are able to annotate the  
631 transcriptome but lack the curation functionalities that are as important as the primary homology-based  
632 identification of candidate genes.

633 As mentioned previously we developed a pipeline called MassBlast (Veríssimo et al. *submitted*), to improve  
634 and accelerate the annotation procedure of RNA-seq data. Besides allowing for operators to query multiple  
635 databases for performing homology searches, the tool also provides a set of parameters that can be defined in  
636 a user-friendly way. MassBlast seeks the best alignment between the sequences in a dataset consisting of  
637 contigs obtained from RNA-seq assemblies and a database of known sequences. In the context of BacHBerry  
638 and as a general validation of the method, MassBlast was used for the analysis of the blackberry  
639 transcriptome, and preliminary results showed a significant improvement in annotation speed (five minutes  
640 versus at least 26 working days, to annotate all known CYP families in blackberry) with very high  
641 accuracies. MassBlast is freely available at <https://github.com/averissimo/mass-blast/>.

### 642 *Sequence design for heterologous expression*

643 Synthetic biology aims at constructing microorganisms capable of producing economically-relevant  
644 chemical products in a cost-effective way, which involves optimizing the functioning of metabolic pathways  
645 and reactions. These pathways and reactions involve proteins that are frequently heterologous. In order to  
646 increase the production of metabolites, it is possible to optimize heterologous protein production in the target  
647 host. This may be done by altering the codons in the gene that encodes for that protein. Each codon is a  
648 sequence of three nucleotides encoding one amino acid that is part of a protein. Since each amino acid can be  
649 encoded by several synonymous codons, a given protein can be expressed using a large number of different  
650 codon combinations.

651 To address this problem, we worked on improving the machine learning models capable of predicting protein  
652 expression levels based on their codon usage frequency (Fernandes et al. 2016) by using Support Vector  
653 Regression (SVR) and Partial Least Squares (PLS). SVR yields predictions that surpass those of PLS. It has  
654 been shown that it is possible to improve the predictive ability of the models by using two additional input  
655 features, namely codon identification number and codon count, besides the already used codon bias and  
656 minimum free energy. In addition, applying ensemble averaging to the SVR or the PLS model improved the  
657 results even further. Different ensembles and features can be tested with the aim of further improving the  
658 prediction models.

#### 659 *Metabolic network optimization*

660 Another goal of the computational activities of BacHBerry was to explore methods for *in silico* metabolic  
661 engineering of phenolic production in *L. lactis* and *C. glutamicum*. For this we selected four model target  
662 compounds: fisetin, *trans*-resveratrol, pelargonidin and quercetin. Additionally, optimization of malonyl-  
663 CoA production was also considered, since this is known to be a major bottleneck in heterologous flavonoid  
664 production. The first step involved performing network optimization for a single population, which included  
665 prediction of knock-out strategies and determination of minimal sets of genetic interventions that could  
666 guarantee an increased production of the target compound production based on stoichiometric models for the  
667 two organisms. These predictions were compared with experimental results and model refinement was

668 conducted in cases of discrepancies. For the second step, we considered a use of microbial consortia for  
669 production purposes. A method was developed that is being used for modeling the production of the  
670 compounds of interest for BacHBerry using *L. lactis* and *C. glutamicum*. Other, possibly larger consortia,  
671 may be considered in the future.

#### 672 Single strain-based production

673 Recently, several computational methods for *in silico* metabolic network optimization have emerged and are  
674 currently being applied in metabolic engineering (Vera et al. 2010; Zomorodi et al. 2012; Machado et al.  
675 2015; Simeonidis and Price 2015; Bergdahl et al. 2015). They are based on different approaches and  
676 rationales, thereby leading to distinct solutions. One key issue is the multitude of results, known to be  
677 dependent on the specific algorithms and solvers, used. In order to address this problem, a consensus-based  
678 approach was developed, following our work on meta-analysis of transcriptomic studies (Caldas et al. 2014).  
679 This method is based on running several optimization procedures and analyzing *a posteriori* the solutions,  
680 looking for a consensus. The rationale is to have rankings of hypotheses that may provide confidence in  
681 particular sets of proposed genetic alterations from various aspects.

682 In the case of a single population of microbes (*i.e.* a monoculture), the genome-scale models of *L. lactis* and  
683 *C. glutamicum* were extended with the biosynthetic pathways of the four above-mentioned target compounds  
684 (**Figure 9**). Subsequently, five different optimization methods were applied to the genome-scale models  
685 (extended with the heterologous pathways): OptKnock (Burgard et al. 2003), MOMA (Segrè et al. 2002),  
686 OptGene (Patil et al. 2005), RoboKod (Stanford et al. 2015), and RobustKnock (Tepper and Shlomi 2010),  
687 and the consensus ranking was obtained for each strain and compound. The hypotheses were generated with  
688 the rank product test and the outputs were lists of gene deletions that would achieve the best ranks using  
689 defined criteria, namely maximum predicted target compound production, maximizing the minimum  
690 predicted target compound production and distance from the wild-type flux distribution. These results were  
691 then further filtered and experimentally validated, resulting in several promising mutant strains where

692 production was increased by up to three-fold. Implementation of this pipeline is possible via a software  
693 package called OptPipe, available at <https://github.com/AndrasHartmann/OptPipe>.

694 We then explored the concept of multi-objective optimization in the field of metabolic engineering when  
695 both continuous and integer decision variables were involved in the model. In particular, we proposed multi-  
696 objective models to suggest reaction deletion strategies and to deal with problems where several functions  
697 must be optimized simultaneously, such as maximization of bioproducts while minimizing toxicity. We thus  
698 introduced Multi-Objective Mixed integer Optimization for metabolic engineering, a computational  
699 framework that aims to model and solve optimization problems, generated for predicting reaction knockout  
700 strategies by means of multi-objective programming. We compared the results from our method with those  
701 obtained by using the well-known bi-level optimization model (OptKnock). Furthermore, we studied two  
702 multi-objective optimization problems arising from the metabolic engineering of microorganisms, showing  
703 that indeed a multi-objective setting is promising, as it expands the set of generated hypotheses and allows  
704 inclusion of multiple goals, simultaneously.

#### 705 For a community of microbes

706 Pure microbial cultures have long been used for production of high-value compounds, as illustrated by the  
707 example of *S. cerevisiae* producing artemisinic acid, an important precursor for the anti-malaria drug  
708 artemisinin, on a large scale (Ro et al. 2006; Lenihan et al. 2008). However, there is increasing evidence that  
709 such processes could run more efficiently if a community of microbial species is used (Logan and Rabaey  
710 2012; Jones et al. 2016; Wen et al. 2016). There are multiple reasons including an increased production  
711 efficiency of the system due to lessening of metabolic burden for individual cells. Another important  
712 consideration is that such community systems might have higher chances of avoiding toxic effects specific  
713 for individual microorganisms. Microbial consortia are believed to be able to sustain more complex  
714 production pathways, as well as being more robust as a group. The challenge, however, remains in  
715 establishing precisely which set of strains or species is best for the production of any given compound.

716 As part of BacHBerry, we introduced an initial topological model and a combinatorial algorithm that enabled  
717 us to propose optimal consortia to produce compounds that were either exogenous to the consortium, or were  
718 endogenous but where interaction between the species in the consortium could improve the production line  
719 (Julien-Laferrière et al. 2016). For initial validations of the model and the method, we applied it to two case-  
720 studies taken from the literature. The first case involved production of two bioactive compounds, penicillin  
721 and cephalosporin C, both of which are antibiotics used in the pharma industry. A synthetic consortium  
722 composed of four different species, three Actinobacteria (*Streptomyces cattleya*, *Rhodococcus jostii* RAH\_1  
723 and *Rc. erythropolis* BG43) and one methanogenic Archaea (*Methanosarcina barkeri*), was then run through  
724 the algorithm, allowing introduction of genetic manipulations and regulatory processes. We showed that the  
725 best solution involved only two of the four species, namely *St. cattleya* and *M. barkeri* (Julien-Laferrière et  
726 al. 2016). Another considered case was a consortium that may be seen as artificial in the sense that it  
727 associates two wild-type populations of microbes, pathogenic  $\gamma$ -proteobacteria *Klebsiella pneumoniae* and  
728 methanogenic Archaea *M. mazei*, which normally do not interact in nature. *K. pneumoniae* is a natural  
729 producer of 1,3-propanediol, which is an important building block for polymer production. Our results  
730 demonstrated that co-cultivation of *K. pneumoniae* with *M. mazei* may enable obtaining higher yields of this  
731 compound in *Klebsiella* (Julien-Laferrière et al. 2016). This supported a previous study involving co-  
732 cultures of *Clostridium sp.* with methanogenes such as *Methanosarcina sp.* CHTI55 (Koesnandar et al.  
733 1990). The developed algorithms are currently being used for developing consortia of polyphenol-producing  
734 Gram-positive bacteria. This task was inspired by the work of Jones and colleagues, who were able to  
735 achieve an overall 970-fold improvement in titers for flavan-3-ols production from phenylpropanoic acids by  
736 using a co-culture of two *E. coli* strains with further optimization of cultivation conditions, as compared to  
737 previous studies that used monocultures for this purpose (Jones et al. 2016).

## 738 **Recovery and purification of polyphenols from fermentation broth**

739 Downstream processing is an important part of any production process and consists of steps required for  
740 recovery and purification of biosynthetic products from fermentation broth, which may have a significant  
741 impact on the final costs of a product (Goldberg 1997). In the design of downstream processing within

742 BacHBerry, *trans*-resveratrol (**Figure 7**) was used as the test product, being one of the model phenolic  
743 compounds that was successfully produced in both *C. glutamicum* and *L. lactis*.

744 In a typical fermentation broth, most of the components present are either hydrophilic (*e.g.*, salts, organic  
745 acids, non-consumed carbon sources (usually glucose), nitrogen source (ammonium sulfate, for instance)) or  
746 are much larger molecules than *trans*-resveratrol (*e.g.*, proteins, nucleic acids, cells in suspension). This  
747 suggested a downstream processing design strategy that used either solubility and hydrophobicity as the  
748 driving forces, or alternatively – size exclusion. Since, at low concentrations (around 100 mg l<sup>-1</sup>),  
749 crystallization would involve high energy costs per kg of product recovered, this option was discarded. Two  
750 possible extraction techniques, namely *liquid-liquid extraction* and *solid-liquid adsorption*, were investigated  
751 as means of recovering and purifying extracellularly produced *trans*-resveratrol from a *C. glutamicum*  
752 fermentation broth containing spent CGXII medium and glucose, due to the higher production efficiency  
753 found in this bacterium. The downstream process design aimed at a recovery of at least 80% and a purity of  
754 95% *trans*-resveratrol.

755 *Liquid-liquid extraction* is a technique to separate a given compound from a feed solution using an  
756 immiscible auxiliary phase that should be selective for that compound (Cox and Rydberg 2004). The  
757 logarithm (base 10) of the partition coefficient of *trans*-resveratrol between the aqueous and different  
758 potential organic phases was determined experimentally for several different solvents (**Figure 10**). The spent  
759 fermentation broth from *C. glutamicum* was used as the aqueous phase in these experiments.

760 All tested organic solvents performed very well in extracting *trans*-resveratrol (minimum log P value is ~0.5,  
761 corresponding to a partition coefficient of 3.2), identifying them as suitable candidates for the recovery of  
762 this compound (when having different impurities present) (**Figure 10**).

763 One of the drawbacks of using liquid-liquid extraction, however, is that it may interfere with protein folding  
764 and stability (Prince and Smith 2004). Besides measuring the partition of the polyphenols of interest in the  
765 aqueous and organic phases, the same analysis was performed for the total protein content. Although a  
766 considerable part of the total protein remained in the aqueous phase, some of the protein seemed to be



767 present in between the two liquid phases (probably as a result of denaturation and foaming) (**Figure 11**).

768 This phenomenon might reduce the recovery of the desired phenolics, as it was also observed experimentally

769 (data not shown).

770 *Solid-liquid adsorption* using hydrophobic resins is another way of purifying *trans*-resveratrol. As in liquid-

771 liquid extraction, the high hydrophobicity of *trans*-resveratrol can be used to recover it selectively from the

772 fermentation broth, leaving the more hydrophilic impurities behind. After an initial resin screen, one of the

773 resins selected was the polystyrene-divinylbenzene based Amberlite XAD-16 from Sigma-Aldrich (particle

774 diameter: 250-841  $\mu\text{m}$ ; dipole moment: 0.3; surface area: 900  $\text{m}^2 \text{g}^{-1}$ ; pore diameter: 100  $\text{\AA}$ ).

775 To test the ability of this resin to recover *trans*-resveratrol from *C. glutamicum* fermentation broth

776 (containing CGXII spent medium and glucose), a lab-scale experiment was performed with an AKTA

777 Explorer chromatographic workstation (GE Healthcare, Sweden) (see **Figure 12**). A 30 ml column was

778 packed with resin and pre-equilibrated with MilliQ water. Afterwards, 10 ml of sample was injected and a

779 washing step with pure water was performed for 10 column volumes (CV). During elution, a gradient of

780 water/ethanol was run for about 12 CV and the peak of *trans*-resveratrol showed up at around 18 CV, at an

781 ethanol percentage of 50% (v/v). Regeneration was performed with 100% ethanol.

782 More hydrophilic molecules such as organic acids, sugars and salts appeared in the flow-through (the first 10

783 CV). During the water/ethanol gradient, molecules desorbed as a function of their hydrophobicity. More

784 hydrophobic molecules, such as *trans*-resveratrol, are supposed to elute at higher ethanol content in the

785 mobile phase. In this case, *trans*-resveratrol appeared to be the compound eluting between 16 and 20 CV,

786 since it also showed absorption at 304 nm (Camont et al. 2009).

787 Although both liquid-liquid extraction and solid-liquid adsorption present their own advantages and

788 disadvantages, a strong point for using adsorption is the use of ethanol instead of any other organic solvent

789 that might not be regarded as GRAS. Moreover, not using liquid-liquid extraction avoids any possible

790 problems related with protein precipitation and emulsification. The ethanol used during elution in adsorption

791 needs to be recovered from water by distillation (the azeotrope poses no problem as 80% (v/v) ethanol can be  
792 used).

793 The proposed downstream process started with a solid-liquid separation step by centrifugation to separate  
794 biomass from the liquid broth. The clarified liquid phase was then sent continuously to one of two adsorption  
795 columns, operating staggered in time. The fraction rich in *trans*-resveratrol was subsequently evaporated,  
796 resulting in a wet slurry containing the polyphenol in the solid phase. This slurry was filtered in a rotary  
797 vacuum filter and washed with water to remove remaining impurities adhered to the surface of the crystals.  
798 The final step was a drying step to obtain the product in its final solid form. All the outlet streams containing  
799 ethanol were directed to a distillation column in order to recover the solvent. A make-up stream ethanol was  
800 added to compensate for losses.

#### 801 **Optimization of phenolic production by fermentation and scale-up of bioprocesses**

802 There are multiple differences between microbial cultivation in shake flasks versus in bioreactors. The  
803 former lacks monitoring and control of various parameters, such as dissolved O<sub>2</sub> concentration and pH (Zelle  
804 et al. 2010). Furthermore, the fed-batch fermentation mode, where one or more nutrient is being supplied  
805 throughout cultivation (Moulton 2014), can be done in a controllable and reproducible fashion only in  
806 bioreactors. These properties make bioreactors a suitable stage for optimization of cultivation conditions,  
807 often leading to significant improvements in production yields (Hujanen et al. 2001; Ratnam et al. 2003; Li  
808 et al. 2010; Rani and Appaiah 2011).

809 From the beginning of the project it was clear that optimization of fermenter operation would need to be  
810 carried out using a model compound, because any newly-identified compounds would only be available at  
811 the end of the project. Hence this section of the project was focused on optimization of production of the  
812 model compound *trans*-resveratrol. However, it was expected that the strategies defined and the conclusions  
813 extracted from these results on *trans*-resveratrol will be valid for the optimization of the fermentation  
814 processes for the production of other target compound. Several *L. lactis* and *C. glutamicum* strains designed  
815 within the project were evaluated and used for fermenter process optimization. The experiments were carried

816 out in agitated 2 to 5 litre fermenters with automatically controlled pH, dissolved oxygen and temperature.  
817 Whenever required, oxygen was transferred to the medium by air injection at the bottom of the vessel.

818 A major goal of the project was the development of an economically feasible and sustainable industrial  
819 process, therefore the following key optimization criteria were defined: (a) minimisation of the operation and  
820 investment cost and (b) maximisation of the production rate and operation yield. To achieve these  
821 optimization goals, the following variables have been investigated: (i) fermentation medium composition, (ii)  
822 dissolved oxygen concentration, (iii) operation strategy (batch vs. fed-batch and integrated vs. separate  
823 downstream operation). Our experiments suggested that there were several constraints limiting the operation  
824 strategy to be implemented, namely: cell metabolism, the stability of *trans*-resveratrol inside the fermenter,  
825 and the affinity of *trans*-resveratrol for the adsorption medium and solvents compared to other compounds  
826 present in the fermentation broth.

827 Medium composition is an important parameter that often has a significant contribution towards the final  
828 production costs. According to our initial experiments, supplementation of *C. glutamicum* minimal medium  
829 with Brain-Heart Infusion (BHI) and a fatty-acid biosynthesis inhibitor, cerulenin, had a strong positive  
830 effect on the final titers of *trans*-resveratrol. However, removing these two components from the medium  
831 reduced medium costs by approximately 300-fold, resulting in a more than 50-fold drop in production costs  
832 (267 €/g compared to 5 €/g of *trans*-resveratrol without supplementation of BHI and cerulenin).

833 We also evaluated the feasibility of producing phenolic compounds from waste materials, namely cheese  
834 whey using *L. lactis* as a host. Whey is a lactose-rich dairy industry by-product with a high environmental  
835 impact, if dumped untreated (Palmieri et al. 2017). To achieve this task, our production strain was modified  
836 to allow it to grow and produce *trans*-resveratrol from lactose by introducing a plasmid carrying the *lac*  
837 operon (Maeda and Gasson 1986). Production at 2 litre-scale was tested using lactose as carbon source and  
838 comparing the productivity between the cultures with glucose or lactose. These experiments demonstrated  
839 that the production of *trans*-resveratrol from lactose, and therefore from whey, is possible with slightly  
840 inferior productivity compared to glucose. However, if whey is chosen as the fermentation medium, other,

841 more important, issues arise. The cost of whey is comparable to the cost of glucose but the cost of the  
842 fermentation medium is only a small part of the total production costs, so variables such as whey availability  
843 and quality, as well as the stability of supply and collecting logistics will determine the feasibility of the  
844 operation more than the small losses in productivity. Clearly, if a phenolic production industrial unit were to  
845 be implemented inside a dairy company with excess whey, the whey might then be the best raw material.

846 Another parameter investigated was the influence of dissolved oxygen concentration on the production  
847 efficiency of *L. lactis*. This bacterium can activate its respiratory chain in the presence of hemin in the  
848 culture medium, and this has been demonstrated to enhance its survival allowing it to grow to higher cell  
849 densities (Duwat et al. 2001). Therefore the production strain efficiency was tested under the following  
850 conditions: aerobic with hemin added to the fermentation broth, aerobic without hemin, semi-aerobic, and  
851 fully anaerobic. In line with the available data, there was a significant increase in the final biomass when the  
852 strain was grown under aerobic conditions with hemin. However, the opposite was observed for the  
853 production of *trans*-resveratrol where the highest amounts of the compound were detected under anaerobic  
854 conditions.

855 Lastly, we have also worked on establishing a system for continuous product removal from the fermentation  
856 broth in order to simplify downstream processing and reduce *trans*-resveratrol oxidation due to prolonged  
857 incubation. Two strategies for continuous adsorption and storage of secreted *trans*-resveratrol in a resin have  
858 been devised and tested: (i) “*in situ*” product removal that consisted of adding a resin directly to the  
859 fermentation broth and (ii) “continuous product removal” (ConPR) that involved constant pumping of the  
860 fermentation broth through a column filled with the adsorption resin. The supernatant was separated from  
861 cells via a tangential flow micro-filtration device, sent through the column, and then returned back into the  
862 fermenter (**Figure 13**).

863 The obtained results demonstrated that both the “*in situ*” and ConPR strategies lowered the *trans*-resveratrol  
864 productivity, probably due to co-adsorption of *p*-coumaric acid, the precursor for *trans*-resveratrol. This  
865 effect was most pronounced in the “*in situ*” operation mode due to continuous contact of the resin with the

866 supernatant and cells. In the ConPR operation mode the effect was reduced due to low supernatant flow rate  
867 through the resin, however adsorption of *trans*-resveratrol was also reduced. In contrast, in the “*in situ*”  
868 mode the amount of adsorbed end-product was nearly identical to the amount of extracellular and  
869 intracellular *trans*-resveratrol produced in control experiments without addition of resin. Although the two  
870 techniques did not increase the overall productivity, they still simplified the downstream processing and may  
871 lead to a production cost reduction by eliminating the need to break down the cells. However, future research  
872 may provide a solution to avoid precursor co-adsorption resulting in increased polyphenol productivity.

### 873 **Dissemination, training, societal and ethical issues**

874 In addition to applying the latest technological advances from life sciences for establishing a pipeline for  
875 microbial-based polyphenol production, BacHBerry analyzed the economic potential of the possible products  
876 developed from this research. The assessment of economic issues of the project has been focused on the  
877 market potential of the natural colorants identified during the screening of the germplasm collection. This  
878 was done along with obtaining an overview of the relevant regulations and procedures for approval of novel  
879 food additives, including colorants, in the EU, the United States (US), and China. Regulations within both  
880 the EU and China confer a strict set of rules for bringing a novel food colorant, even a natural one, to the  
881 market. These rules consider not only the composition and purity of the substance, but also the process  
882 through which the compounds have been produced. In contrast, in the US, a new food additive can gain  
883 access to the market by simply passing the GRAS assessment. Hence, if a novel polyphenolic colorant could  
884 be produced using engineered microbial cell factories, the US market should probably be the first market to  
885 be targeted, owing to less strict regulatory requirements for the approval.

886 Another topic addressed in BacHBerry was access and benefit sharing of genetic resources. As the project  
887 was built upon screening of berry species from multiple countries, including China, Russia, Chile, Portugal,  
888 and the UK, it is important to harness these resources in a responsible manner. *The Nagoya Protocol on*  
889 *Access to Genetic Resources and the Fair and Equitable Sharing of Benefits Arising from their Utilization*  
890 (NP-ABS) is an agreement consisting of a set of rules for fair sharing of monetary and non-monetary benefits

891 arising from the usage and commercialization of genetic resources and traditional knowledge associated with  
892 those. It aims at achieving fairness and equity of gain distribution, especially towards indigenous and local  
893 communities (SCBD 2011; Morgera et al. 2014). The Protocol came into force in the EU as of November  
894 20<sup>th</sup> 2016. As most of the BacHBerry project partners were from EU Member States, we began collecting  
895 information about the best practices suggested from both the Convention on Biological Diversity (CBD) and  
896 the EU, and kept monitoring the relevant guidelines and updates from the EU on implementing ABS. Among  
897 the home countries of the project partners China, Denmark, France, Germany, Switzerland, and UK ratified  
898 or accepted the NP-ABS (as of November 2016) and began implementing the aforementioned regulations  
899 (European Parliament 2014). The UK also made its own regulation on implementing the Protocol (European  
900 Parliament 2014; NMRO (UK) 2015; Secretary of State (UK) 2015).

901 Funded under the theme of “Knowledge-based Bio-Economy”, a final aim of the project was to provide  
902 scientific and technical solutions to harness berry resources, as well as to contribute to sustainable  
903 development of the economy. We consequently investigated the possible contributions of our research to  
904 build a more sustainable society based on environmental, economical, and social impacts (Pei and Schmidt  
905 2016). Furthermore, we conducted several open dialog events, namely workshops and interviews with  
906 potential stakeholders from Austria, Denmark, and China. One of the surprising outcomes was that the  
907 general public raised concerns over the use of genetically-modified organisms (GMOs) in the pipeline a lot  
908 less frequently than originally expected. A free, science-based game for mobile devices called “Berry  
909 Maker” was produced (<http://www.berrymaker.com>). The game allows the user to follow the pipeline  
910 designed within our project, from acquisition of diverse berries, to extraction and analysis of biosynthetic  
911 genes, to microbial fermentation-based production of specific polyphenols and running a biotech factory that  
912 is exposed to real world situations, such as price fluctuations. The game is also available on iTunes<sup>1</sup> and  
913 Google Play<sup>2</sup> store. Lastly, a 22 minute documentary film was produced, showing sampling of berries,

---

<sup>1</sup> See: <https://itunes.apple.com/us/app/berrymaker-dna-soda-factory/id1163115116?ls=1&mt=8>

<sup>2</sup> See: <https://play.google.com/store/apps/details?id=com.biofaction.berrymaker>

914 analysis of interesting compounds, and experiments with natural color pigments for food products from  
915 across Europe (see: <https://vimeo.com/193467652>).

## 916 **Conclusions**

917 During three years of the project, the BacHBerry partners successfully established a complete pipeline for  
918 screening, bioprospecting, identification, and microbial production of plant phenolic compounds in Gram-  
919 positive bacteria. A summary of the most important results is presented in **Table 5**. The major highlights are  
920 construction of the extensive germplasm collection database, establishment of a method for high-throughput  
921 screening of bioactive extracts using the yeast-based human disease models, generation of a vast berry  
922 transcriptome database, construction of microbial cell factories in *L. lactis* and *C. glutamicum* for the  
923 production of polyphenols with yields comparable to those currently available from the literature, release of  
924 several software packages for transcriptome analysis and *in silico*-based metabolic engineering, as well as  
925 optimization of fermentation conditions and downstream processing for the selected organisms aimed at  
926 increasing the yield and the purify of the final products. The pipeline has been completely validated and is  
927 available to be picked up by the biotech industry for further improvement and commercialization. The  
928 developed methods and protocols should be suitable for discovery, as well as production of most groups of  
929 polyphenols in the above-mentioned organisms, without a need for significant alterations and adaptations.  
930 Overall, this project provided the scientific community with enabling tools and resources to further  
931 advancing research on berries, phenolic compounds, and metabolic engineering of microorganisms.  
932 Additional efforts were also focused on communication with stakeholders, public outreach, and  
933 popularization of science among the younger generation. The majority of the project outcomes is already  
934 available to the scientific community and the industry in form of peer-reviewed articles, or will be published  
935 within the coming year.

936

## 937 **Acknowledgments**

938 The authors would like to thank the European Union's Seventh Framework Programme (BacHBerry, Project  
 939 No. FP7-613793, and FP7-PEOPLE-2013-COFUND, Project No. FP7-609405) for the financial support.  
 940 AD, RF, MHHN, PG, SGS, and JF would also like to acknowledge the Novo Nordisk Foundation. We  
 941 express our gratitude to Dr. Martha Cyert (Stanford School of Medicine, EUA), Dr. Hitoshi Shimoi (National  
 942 Research Institute of Brewing, Japan) and Dr. Yoshio Araki (Graduate School of Biosphere Science,  
 943 Hiroshima University, Japan) for providing the yeast strain YAA5. We also thank Dr. Ian Macraedie, RMIT  
 944 University, Australia) for providing the plasmid p416\_GPDpr-GFP-A42.

945

## 946 **Abbreviations**

947 4CL, 4-coumaryl-CoA ligase; AD, Alzheimer's Disease; ANS, anthocyanidin synthase; ALS, amyotrophic  
 948 lateral sclerosis; anthocyanidin synthase; ANR, anthocyanidin reductase; C3'H, *p*-coumaroyl  
 949 shikimate/quinate 3'-hydroxylase; C4H, cinnamate 4-hydroxylase; CBD, Convention on Biological  
 950 Diversity; CaN, calcineurin; CDRE, Calcineurin-Dependent Responsive Element; CHI, chalcone isomerase;  
 951 CHR, chalcone reductase; CHS, chalcone synthase; CPR, cytochrome-P450 reductase; CVs, column  
 952 volumes; CYP, cytochromes P450; DFR, dihydroflavonol 4-reductase; F3H, flavanone 3-hydroxylase; F3'H,  
 953 flavonoid 3'-hydroxylase; F3'5'H, flavonoid 3'-5'-hydroxylase; FLS, flavonol synthase; FNS, flavone  
 954 synthase; FP7, European Commission's the Seventh Framework programme; GFP, Green Fluorescent  
 955 Protein; GRAS, generally regarded as safe; HCT, hydroxycinnamoyl-CoA shikimate/quinate  
 956 hydroxycinnamoyltransferase; HD, Huntington's Disease; HPLC, high-performance liquid chromatography;  
 957 LAR, leucoanthocyanidin reductase; LC-MS, liquid-chromatography mass-spectrometry; NDs,  
 958 Neurodegenerative diseases; NFAT, Nuclear Factor of Activated T-cells; NFκB, Nuclear Factor κB; NP-  
 959 ABS, The Nagoya protocol on access to genetic resources and the fair and equitable sharing of benefits  
 960 arising from their utilization; OMT, *O*-methyltransferase; PAL, phenylalanine ammonia-lyase; PCA,  
 961 principal component analysis; PD, Parkinson's Disease; PLS, partial least squares; RNA-seq, RNA



962 sequencing technology; STS, stilbene synthase; SVR, support vector regression; TAL, tyrosine ammonia-  
963 lyase; UFGT, flavonoid 3-*O*-glucosyltransferase.

964

## 965 **Appendices**

966 Supplementary\_Materials.docx

967 Supplementary\_Tables.docx

968

## 969 **References**

970 Andrade R, Wannagat M, Klein CC, et al (2016) Enumeration of minimal stoichiometric precursor sets in  
971 metabolic networks. *Algorithms Mol Biol* 11:25. doi: 10.1186/s13015-016-0087-3

972 Aranaz I, Acosta N, Mengibar MN, et al (2016) InFiQuS: Making the Best of Leftovers. In: Domínguez de  
973 María P (ed) *Industrial biorenewables : a practical viewpoint*. John Wiley & Sons, Inc., Hoboken, US,  
974 pp 363–364

975 Bak S, Olsen CE, Halkier BA, Møller BL (2000) Transgenic tobacco and *Arabidopsis* plants expressing the  
976 two multifunctional sorghum cytochrome P450 enzymes, CYP79A1 and CYP71E1, are cyanogenic and  
977 accumulate metabolites derived from intermediates in Dhurrin biosynthesis. *Plant Physiol* 123:1437–  
978 48. doi: 10.1104/pp.123.4.1437

979 Becker J, Wittmann C (2012) Bio-based production of chemicals, materials and fuels – *Corynebacterium*  
980 *glutamicum* as versatile cell factory. *Curr Opin Biotechnol* 23:631–640. doi:  
981 10.1016/j.copbio.2011.11.012

982 Becker J, Armstrong G, Vandermerwe M, et al (2003) Metabolic engineering of for the synthesis of the

983 wine-related antioxidant resveratrol. FEMS Yeast Res 4:79–85. doi: 10.1016/S1567-1356(03)00157-0

984 Beekwilder J, Wolswinkel R, Jonker H, et al (2006) Production of resveratrol in recombinant  
 985 microorganisms. Appl Environ Microbiol 72:5670–5672. doi: 10.1128/AEM.00609-06

986 Bergdahl B, Sonnenschein N, Machado D, et al (2015) Genome-Scale Models. In: Villadsen J (ed)  
 987 Fundamental Bioengineering. Wiley-VCH Verlag GmbH & Co. KGaA, Weinheim, Germany, pp 143–  
 988 182

989 Bhan N, Xu P, Khalidi O, Koffas MAG (2013) Redirecting carbon flux into malonyl-CoA to improve  
 990 resveratrol titers: Proof of concept for genetic interventions predicted by OptForce computational  
 991 framework. Chem Eng Sci 103:109–114. doi: 10.1016/j.ces.2012.10.009

992 Bharadwaj P, Martins R, Macreadie I (2010) Yeast as a model for studying Alzheimer’s disease. FEMS  
 993 Yeast Res 10:961–969. doi: 10.1111/j.1567-1364.2010.00658.x

994 Blomstedt CK, Gleadow RM, O’Donnell N, et al (2012) A combined biochemical screen and TILLING  
 995 approach identifies mutations in *Sorghum bicolor* L. Moench resulting in acyanogenic forage  
 996 production. Plant Biotechnol J 10:54–66. doi: 10.1111/j.1467-7652.2011.00646.x

997 Bodvard K, Jörhov A, Blomberg A, et al (2013) The Yeast Transcription Factor Crz1 Is Activated by Light  
 998 in a Ca<sup>2+</sup>/Calcineurin-Dependent and PKA-Independent Manner. PLoS One 8:e53404. doi:  
 999 10.1371/journal.pone.0053404

1000 Borrill P, Ramirez-Gonzalez R, Uauy C (2016) expVIP: a Customizable RNA-seq Data Analysis and  
 1001 Visualization Platform. Plant Physiol 170:2172–86. doi: 10.1104/pp.15.01667

1002 Burgard AP, Pharkya P, Maranas CD (2003) Optknock: A bilevel programming framework for identifying  
 1003 gene knockout strategies for microbial strain optimization. Biotechnol Bioeng 84:647–657. doi:  
 1004 10.1002/bit.10803

1005 Burkovski A (ed) (2008) *Corynebacteria: Genomics and Molecular Biology*. Caister Academic Press

1006 Caldas J, Vinga S, Nordmann A, et al (2014) Global Meta-Analysis of Transcriptomics Studies. *PLoS One*  
 1007 9:e89318. doi: 10.1371/journal.pone.0089318

1008 Camont L, Cottart C-H, Rhayem Y, et al (2009) Simple spectrophotometric assessment of the trans-/cis-  
 1009 resveratrol ratio in aqueous solutions. *Anal Chim Acta* 634:121–128. doi: 10.1016/j.aca.2008.12.003

1010 Cao W, Ma W, Zhang B, et al (2016) Improved pinocembrin production in *Escherichia coli* by engineering  
 1011 fatty acid synthesis. *J Ind Microbiol Biotechnol* 43:557–566. doi: 10.1007/s10295-015-1725-3

1012 Carvalho E, Franceschi P, Feller A, et al (2013) A targeted metabolomics approach to understand differences  
 1013 in flavonoid biosynthesis in red and yellow raspberries. *Plant Physiol Biochem PPB* 72:79–86. doi:  
 1014 10.1016/j.plaphy.2013.04.001

1015 Catalgol B, Batirel S, Taga Y, Ozer NK (2012) Resveratrol: French Paradox Revisited. *Front Pharmacol*  
 1016 3:141. doi: 10.3389/fphar.2012.00141

1017 Chaovanalikit A, Thompson MM, Wrolstad RE (2004) Characterization and Quantification of Anthocyanins  
 1018 and Polyphenolics in Blue Honeysuckle (*Lonicera caerulea* L.). *J Agric Food Chem* 52:848–852. doi:  
 1019 10.1021/JF030509O

1020 Chen X, Zhou L, Tian K, et al (2013) Metabolic engineering of *Escherichia coli*: A sustainable industrial  
 1021 platform for bio-based chemical production. *Biotechnol Adv* 31:1200–1223. doi:  
 1022 10.1016/j.biotechadv.2013.02.009

1023 Choi O, Wu C-Z, Kang SY, et al (2011) Biosynthesis of plant-specific phenylpropanoids by construction of  
 1024 an artificial biosynthetic pathway in *Escherichia coli*. *J Ind Microbiol Biotechnol* 38:1657–1665. doi:  
 1025 10.1007/s10295-011-0954-3

1026 Costa RS, Hartmann A, Vinga S (2016) Kinetic modeling of cell metabolism for microbial production. *J*

- 1027 Biotechnol 219:126–141. doi: 10.1016/j.jbiotec.2015.12.023
- 1028 Costa RS, Veríssimo A, Vinga S, et al (2014) Ki MoSys: a web-based repository of experimental data for  
1029 KInetic MOdels of biological SYStems. BMC Syst Biol 8:85. doi: 10.1186/s12918-014-0085-3
- 1030 Cox M, Rydberg J (2004) Introduction to Solvent Extraction. In: Rydberg J, Cox M, Musikas C, Choppin  
1031 GR (eds) Solvent extraction principles and practice, 2nd edn. Marcel Decker, Inc., New York, USA,
- 1032 Cragg GM, Newman DJ (2013) Natural products: A continuing source of novel drug leads. Biochim Biophys  
1033 Acta - Gen Subj 1830:3670–3695. doi: 10.1016/j.bbagen.2013.02.008
- 1034 Dobson R, Gray V, Rumbold K (2012) Microbial utilization of crude glycerol for the production of value-  
1035 added products. J Ind Microbiol Biotechnol 39:217–226. doi: 10.1007/s10295-011-1038-0
- 1036 Duarte K, Rocha-Santos TAP, Freitas AC, Duarte AC (2012) Analytical techniques for discovery of  
1037 bioactive compounds from marine fungi. TrAC Trends Anal Chem 34:97–110. doi:  
1038 10.1016/j.trac.2011.10.014
- 1039 Dumont FJ (2000) FK506, an immunosuppressant targeting calcineurin function. Curr Med Chem 7:731–48.
- 1040 Duwat P, Sourice S, Cesselin B, et al (2001) Respiration capacity of the fermenting bacterium *Lactococcus*  
1041 *lactis* and its positive effects on growth and survival. J Bacteriol 183:4509–16. doi:  
1042 10.1128/JB.183.15.4509-4516.2001
- 1043 Dvora H, Koffas MAG (2013) Microbial production of flavonoids and terpenoids. In: McNeil B, Archer D,  
1044 Giavasis I, Harvey L (eds) Microbial Production of Food Ingredients, Enzymes and Nutraceuticals.  
1045 Woodhead Publishing, Oxford, UK, pp 234–261
- 1046 Eggeling L, Bott M (eds) (2005) Handbook of *Corynebacterium glutamicum*. CRC Press, Boca Raton, US
- 1047 European Parliament (2014) REGULATION (EU) No 511/2014 OF THE EUROPEAN PARLIAMENT  
1048 AND OF THE COUNCIL of 16 April 2014 on compliance measures for users from the Nagoya

1049 Protocol on Access to Genetic Resources and the Fair and Equitable Sharing of Benefits Arising from  
 1050 their Utilizati.

1051 Falcone Ferreyra ML, Rius SP, Casati P (2012) Flavonoids: biosynthesis, biological functions, and  
 1052 biotechnological applications. *Front Plant Sci* 3:222. doi: 10.3389/fpls.2012.00222

1053 Fernandes A, Vinga S, Kern A, et al (2016) Improving Protein Expression Prediction Using Extra Features  
 1054 and Ensemble Averaging. *PLoS One* 11:e0150369. doi: 10.1371/journal.pone.0150369

1055 Figueira I, Menezes R, Macedo D, et al Polyphenols beyond barriers: a glimpse into the brain.

1056 Fredes C, Yousef GG, Robert P, et al (2014) Anthocyanin profiling of wild maqui berries ( *Aristotelia*  
 1057 *chilensis* [Mol.] Stuntz) from different geographical regions in Chile. *J Sci Food Agric* 94:2639–2648.  
 1058 doi: 10.1002/jsfa.6602

1059 Frelet-Barrand A, Boutigny S, Moyet L, et al (2010) *Lactococcus lactis*, an Alternative System for  
 1060 Functional Expression of Peripheral and Intrinsic Arabidopsis Membrane Proteins. *PLoS One* 5:e8746.  
 1061 doi: 10.1371/journal.pone.0008746

1062 Funakoshi-Tago M, Nakamura K, Tago K, et al (2011) Anti-inflammatory activity of structurally related  
 1063 flavonoids, Apigenin, Luteolin and Fisetin. *Int Immunopharmacol* 11:1150–1159. doi:  
 1064 10.1016/j.intimp.2011.03.012

1065 Garcia G, Santos CN do, Menezes R (2016) High-Throughput Yeast-Based Reporter Assay to Identify  
 1066 Compounds with Anti-inflammatory Potential. In: Matthiesen R (ed) *Proteostasis*. Springer, New York,  
 1067 USA, pp 441–452

1068 Gaspar P, Carvalho AL, Vinga S, et al (2013) From physiology to systems metabolic engineering for the  
 1069 production of biochemicals by lactic acid bacteria. *Biotechnol Adv* 31:764–88. doi:  
 1070 10.1016/j.biotechadv.2013.03.011

1071 Ghosh D, Scheepens A (2009) Vascular action of polyphenols. *Mol Nutr Food Res* 53:322–331. doi:  
 1072 10.1002/mnfr.200800182

1073 Giusti MM, Wrolstad RE (2003) Acylated anthocyanins from edible sources and their applications in food  
 1074 systems. *Biochem Eng J* 14:217–225. doi: 10.1016/S1369-703X(02)00221-8

1075 Goldberg E (1997) *Handbook of Downstream Processing*. Springer Netherlands, London, UK

1076 Goszcz K, Deakin SJ, Duthie GG, et al (2015) Antioxidants in Cardiovascular Therapy: Panacea or False  
 1077 Hope? *Front Cardiovasc Med* 2:29. doi: 10.3389/fcvm.2015.00029

1078 Grand View Research Inc. (2016) *Polyphenols Market Analysis By Product (Grape Seed, Green Tea,  
 1079 Apple), By Application (Functional Food, Functional Beverages, Dietary Supplements) And Segment  
 1080 Forecasts To 2024*.

1081 Halvorsen BL, Carlsen MH, Phillips KM, et al (2006) Content of redox-active compounds (ie, antioxidants)  
 1082 in foods consumed in the United States. *Am J Clin Nutr* 84:95–135.

1083 Hartmann U, Sagasser M, Mehrtens F, et al (2005) Differential combinatorial interactions of cis-acting  
 1084 elements recognized by R2R3-MYB, BZIP, and BHLH factors control light-responsive and tissue-  
 1085 specific activation of phenylpropanoid biosynthesis genes. *Plant Mol Biol* 57:155–171. doi:  
 1086 10.1007/s11103-004-6910-0

1087 Hartmann A, Lemos JM, Costa RS, et al (2015a) Identification of switched ARX models via convex  
 1088 optimization and expectation maximization. *J Process Control* 28:9–16. doi:  
 1089 10.1016/j.jprocont.2015.02.003

1090 Hartmann A, Lemos JM, Vinga S (2015b) Modeling multiple experiments using regularized optimization: A  
 1091 case study on bacterial glucose utilization dynamics. *Comput Biol Med* 63:301–309. doi:  
 1092 10.1016/j.combiomed.2014.08.027

- 1093 Hartmann A, Neves AR, Lemos JM, Vinga S (2016) Identification and automatic segmentation of  
 1094 multiphasic cell growth using a linear hybrid model. *Math Biosci* 279:83–89. doi:  
 1095 10.1016/j.mbs.2016.06.013
- 1096 Heldt H-W, Piechulla B (2011) *Plant Biochemistry*, 4th edn. Elsevier, London, UK
- 1097 Hellström JK, Törrönen AR, Mattila PH (2009) Proanthocyanidins in Common Food Products of Plant  
 1098 Origin. *J Agric Food Chem* 57:7899–7906. doi: 10.1021/jf901434d
- 1099 Hernández I, Molenaar D, Beekwilder J, et al (2007) Expression of plant flavor genes in *Lactococcus lactis*.  
 1100 *Appl Environ Microbiol* 73:1544–52. doi: 10.1128/AEM.01870-06
- 1101 Herrmann K, Nagel CW (1989) Occurrence and content of hydroxycinnamic and hydroxybenzoic acid  
 1102 compounds in foods. *Crit Rev Food Sci Nutr* 28:315–347. doi: 10.1080/10408398909527504
- 1103 Hujanen M, Linko S, Linko Y-Y, Leisola M (2001) Optimisation of media and cultivation conditions for L  
 1104 (+)(S)-lactic acid production by *Lactobacillus casei* NRRL B-441. *Appl Microbiol Biotechnol*  
 1105 56:126–130. doi: 10.1007/s002530000501
- 1106 Häkkinen SH, Kärenlampi SO, Heinonen IM, et al (1999) Content of the Flavonols Quercetin, Myricetin,  
 1107 and Kaempferol in 25 Edible Berries. *J Agric Food Chem* 47:2274–2279. doi: 10.1021/JF9811065
- 1108 Jain N, Ramawat KG (2013) Nutraceuticals and Antioxidants in Prevention of Diseases. In: Ramawat KG,  
 1109 Merillon J-M (eds) *Natural Products*. Springer Berlin Heidelberg, Berlin, Germany, pp 2559–2580
- 1110 Jash SK, Mondal S (2014) Bioactive flavonoid fisetin – a molecule of pharmacological interest. *Signpost*  
 1111 *Open Access J Org Biomol Chem* 2:89–128.
- 1112 Jeandet P, Bessis R, Gautheron B (1991) The Production of Resveratrol (3,5,4'-trihydroxystilbene) by Grape  
 1113 Berries in Different Developmental Stages. *Am J Enol Vitic* 42:41–46.
- 1114 Jendresen CB, Stahlhut SG, Li M, et al (2015) Highly Active and Specific Tyrosine Ammonia-Lyases from

- 1115 Diverse Origins Enable Enhanced Production of Aromatic Compounds in Bacteria and *Saccharomyces*  
 1116 *cerevisiae*. Appl Environ Microbiol 81:4458–76. doi: 10.1128/AEM.00405-15
- 1117 Jojima T, Inui M, Yukawa H (2013) Biorefinery Applications of *Corynebacterium glutamicum*. In: Yukawa  
 1118 H, Inui M (eds) *Corynebacterium glutamicum: Biology and Biotechnology*. Springer, Berlin, Germany,  
 1119 pp 149–172
- 1120 Jones JA, Vernacchio VR, Sinkoe AL, et al (2016) Experimental and computational optimization of an  
 1121 *Escherichia coli* co-culture for the efficient production of flavonoids. Metab Eng 35:55–63. doi:  
 1122 10.1016/j.ymben.2016.01.006
- 1123 Josuttis M, Verrall S, Stewart D, et al (2013) Genetic and environmental effects on tannin composition in  
 1124 strawberry (*Fragaria* × *ananassa*) cultivars grown in different European locations. J Agric Food Chem  
 1125 61:790–800. doi: 10.1021/jf303725g
- 1126 Ju S, Tardiff DF, Han H, et al (2011) A Yeast Model of FUS/TLS-Dependent Cytotoxicity. PLoS Biol  
 1127 9:e1001052. doi: 10.1371/journal.pbio.1001052
- 1128 Julien-Laferrrière A, Bulteau L, Parrot D, et al (2016) A Combinatorial Algorithm for Microbial Consortia  
 1129 Synthetic Design. Sci Rep 6:29182. doi: 10.1038/srep29182
- 1130 Kallscheuer N, Vogt M, Bott M, Marienhagen J (2017) Functional expression of plant-derived O-  
 1131 methyltransferase, flavanone 3-hydroxylase, and flavonol synthase in *Corynebacterium glutamicum* for  
 1132 production of pterostilbene, kaempferol, and quercetin. J Biotechnol. doi:  
 1133 10.1016/j.jbiotec.2017.01.006
- 1134 Kallscheuer N, Vogt M, Kappelmann J, et al (2016a) Identification of the phd gene cluster responsible for  
 1135 phenylpropanoid utilization in *Corynebacterium glutamicum*. Appl Microbiol Biotechnol 100:1871–  
 1136 1881. doi: 10.1007/s00253-015-7165-1
- 1137 Kallscheuer N, Vogt M, Marienhagen J (2016b) A Novel Synthetic Pathway Enables Microbial Production



1138 of Polyphenols Independent from the Endogenous Aromatic Amino Acid Metabolism. ACS Synth Biol  
 1139 acssynbio.6b00291. doi: 10.1021/acssynbio.6b00291

1140 Kallscheuer N, Vogt M, Stenzel A, et al (2016c) Construction of a *Corynebacterium glutamicum* platform  
 1141 strain for the production of stilbenes and (2S)-flavanones. Metab Eng 38:47–55. doi:  
 1142 10.1016/j.ymben.2016.06.003

1143 Kasiotis KM, Pratsinis H, Kletsas D, Haroutounian SA (2013) Resveratrol and related stilbenes: their anti-  
 1144 aging and anti-angiogenic properties. Food Chem Toxicol 61:112–120. doi: 10.1016/j.fct.2013.03.038

1145 Katz M, Durhuus T, Smits HP, FÖRSTER J (2011) Production of metabolites.

1146 Kim HJ, Chang EJ, Cho SH, et al (2002) Antioxidative Activity of Resveratrol and Its Derivatives Isolated  
 1147 from Seeds of *Paeonia lactiflora*. Biosci Biotechnol Biochem 66:1990–1993. doi: 10.1271/bbb.66.1990

1148 Koesnandar A, Nishio N, Nagai S (1990) Stimulation by cysteine on growth of *Clostridium thermoaceticum*  
 1149 in minimal medium. Appl Microbiol Biotechnol 32:711–714. doi: 10.1007/BF00164746

1150 Kortmann M, Kuhl V, Klaffl S, Bott M (2015) A chromosomally encoded T7 RNA polymerase-dependent  
 1151 gene expression system for *Corynebacterium glutamicum*: construction and comparative evaluation at  
 1152 the single-cell level. Microb Biotechnol 8:253–265. doi: 10.1111/1751-7915.12236

1153 Kritzer JA, Hamamichi S, McCaffery JM, et al (2009) Rapid selection of cyclic peptides that reduce  $\alpha$ -  
 1154 synuclein toxicity in yeast and animal models. Nat Chem Biol 5:655–663. doi: 10.1038/nchembio.193

1155 Krobitsch S, Lindquist S (2000) Aggregation of huntingtin in yeast varies with the length of the  
 1156 polyglutamine expansion and the expression of chaperone proteins. Proc Natl Acad Sci U S A  
 1157 97:1589–94. doi: 10.1073/PNAS.97.4.1589

1158 Krömer JO, Sorgenfrei O, Klopprogge K, et al (2004) In-depth profiling of lysine-producing  
 1159 *Corynebacterium glutamicum* by combined analysis of the transcriptome, metabolome, and fluxome. J

1160 Bacteriol 186:1769–84. doi: 10.1128/JB.186.6.1769-1784.2004

1161 Kuipers OP, Beerthuyzen MM, Siezen RJ, et al (1993) Characterization of the nisin gene cluster  
 1162 *nisABTCIPR* of *Lactococcus lactis*. Requirement of expression of the *nisA* and *nisI* genes for  
 1163 development of immunity. Eur J Biochem 216:281–291. doi: 10.1111/j.1432-1033.1993.tb18143.x

1164 Kuipers OP, de Ruyter PGG., Kleerebezem M, de Vos WM (1998) Quorum sensing-controlled gene  
 1165 expression in lactic acid bacteria. J Biotechnol 64:15–21. doi: 10.1016/S0168-1656(98)00100-X

1166 Kula M, Majdan M, Głód D, Krauze-Baranowska M (2016) Phenolic composition of fruits from different  
 1167 cultivars of red and black raspberries grown in Poland. J Food Compos Anal 52:74–82. doi:  
 1168 10.1016/j.jfca.2016.08.003

1169 Kunji ERS, Chan KW, Slotboom DJ, et al (2005) Eukaryotic membrane protein overproduction in  
 1170 *Lactococcus lactis*. Curr Opin Biotechnol 16:546–551. doi: 10.1016/j.copbio.2005.08.006

1171 Kwiatkowski TJ, Bosco DA, LeClerc AL, et al (2009) Mutations in the FUS/TLS Gene on Chromosome 16  
 1172 Cause Familial Amyotrophic Lateral Sclerosis. Science 323:1205–1208. doi: 10.1126/science.1166066

1173 Kårlund A, Moor U, Sandell M, Karjalainen R (2014) The Impact of Harvesting, Storage and Processing  
 1174 Factors on Health-Promoting Phytochemicals in Berries and Fruits. Processes 2:596–624. doi:  
 1175 10.3390/pr2030596

1176 Lahtinen S, Ouwehand AC, Salminen S, von Wright A (eds) (2011) Lactic Acid Bacteria: Microbiological  
 1177 and Functional Aspects, 4th edn. CRC Press, Boca Raton, USA

1178 Landete JM (2011) Ellagitannins, ellagic acid and their derived metabolites: A review about source,  
 1179 metabolism, functions and health. Food Res Int 44:1150–1160. doi: 10.1016/j.foodres.2011.04.027

1180 Lenihan JR, Tsuruta H, Diola D, et al (2008) Developing an industrial artemisinic acid fermentation process  
 1181 to support the cost-effective production of antimalarial artemisinin-based combination therapies.

1182 Biotechnol Prog 24:1026–1032. doi: 10.1002/btpr.27

1183 Li H, Qiu T, Huang G, et al (2010) Production of gamma-aminobutyric acid by *Lactobacillus brevis*  
 1184 NCL912 using fed-batch fermentation. Microb Cell Fact 9:85. doi: 10.1186/1475-2859-9-85

1185 Li X, Sun H, Pei J, et al (2012) De novo sequencing and comparative analysis of the blueberry transcriptome  
 1186 to discover putative genes related to antioxidants. Gene 511:54–61. doi: 10.1016/j.gene.2012.09.021

1187 Linares DM, Kok J, Poolman B (2010) Genome sequences of *Lactococcus lactis* MG1363 (revised) and  
 1188 NZ9000 and comparative physiological studies. J Bacteriol 192:5806–12. doi: 10.1128/JB.00533-10

1189 Liu Y-B, Long M-X, Yin Y-J, et al (2013) Physiological roles of mycothiol in detoxification and tolerance to  
 1190 multiple poisonous chemicals in *Corynebacterium glutamicum*. Arch Microbiol 195:419–429. doi:  
 1191 10.1007/s00203-013-0889-3

1192 Logan BE, Rabaey K (2012) Conversion of Wastes into Bioelectricity and Chemicals by Using Microbial  
 1193 Electrochemical Technologies. Science 337:686–690. doi: 10.1126/science.1217412

1194 Luo J, Nishiyama Y, Fuell C, et al (2007) Convergent evolution in the BAHD family of acyl transferases:  
 1195 identification and characterization of anthocyanin acyl transferases from *Arabidopsis thaliana*. Plant J  
 1196 50:678–695. doi: 10.1111/j.1365-313X.2007.03079.x

1197 MacDonald MJ, D’Cunha GB (2007) A modern view of phenylalanine ammonia lyase. Biochem Cell Biol  
 1198 85:273–282. doi: 10.1139/O07-018

1199 Machado D, Herrgård MJ, Rocha I (2015) Modeling the Contribution of Allosteric Regulation for Flux  
 1200 Control in the Central Carbon Metabolism of *E. coli*. Front Bioeng Biotechnol 3:154. doi:  
 1201 10.3389/fbioe.2015.00154

1202 Maeda S, Gasson MJ (1986) Cloning, expression and location of the *Streptococcus lactis* gene for phospho-  
 1203  $\beta$ -d-galactosidase. Microbiology 132:331–340. doi: 10.1099/00221287-132-2-331

- 1204 Maher P (2009) Modulation of multiple pathways involved in the maintenance of neuronal function during  
1205 aging by fisetin. *Genes Nutr* 4:297–307. doi: 10.1007/s12263-009-0142-5
- 1206 Maher P, Dargusch R, Ehren JL, et al (2011) Fisetin Lowers Methylglyoxal Dependent Protein Glycation  
1207 and Limits the Complications of Diabetes. *PLoS One* 6:e21226. doi: 10.1371/journal.pone.0021226
- 1208 Manganaris GA, Goulas V, Vicente AR, Terry LA (2014) Berry antioxidants: small fruits providing large  
1209 benefits. *J Sci Food Agric* 94:825–833. doi: 10.1002/jsfa.6432
- 1210 Marhuenda J, Alemán MD, Gironés-Vilaplana A, et al (2016) Phenolic Composition, Antioxidant Activity,  
1211 and In Vitro Availability of Four Different Berries. *J Chem* 2016:1–7. doi: 10.1155/2016/5194901
- 1212 Marienhagen J, Bott M (2013) Metabolic engineering of microorganisms for the synthesis of plant natural  
1213 products. *J Biotechnol* 163:166–178. doi: 10.1016/j.jbiotec.2012.06.001
- 1214 Mazur SP, Sørnsteby A, Nes A, et al (2014) Effects of Post-Flowering Environmental Variation along an  
1215 Altitudinal Gradient on Chemical Composition of “Glen Ample” Red Raspberry (*Rubus idaeus* L.). *Eur*  
1216 *J Hort Sci* 79:267–277.
- 1217 McDougall G, Martinussen I, Stewart D (2008) Towards fruitful metabolomics: High throughput analyses of  
1218 polyphenol composition in berries using direct infusion mass spectrometry. *J Chromatogr B* 871:362–  
1219 369. doi: 10.1016/j.jchromb.2008.06.032
- 1220 Menezes R, Tenreiro S, Macedo D, et al (2015) From the baker to the bedside: yeast models of Parkinson’s  
1221 disease. *Microb Cell* 2:262–279. doi: 10.15698/mic2015.08.219
- 1222 Miladinović B, Kostić M, Šavikin K, et al (2014) Chemical Profile and Antioxidative and Antimicrobial  
1223 Activity of Juices and Extracts of 4 Black Currants Varieties ( *Ribes nigrum* L.). *J Food Sci* 79:C301–  
1224 C309. doi: 10.1111/1750-3841.12364
- 1225 Milivojević J, Maksimović V, Nikolić M, et al (2011) Chemical and antioxidant properties of cultivated and

1226 wild *Fragaria* and *Rubus* berries. J Food Qual 34:1–9. doi: 10.1111/j.1745-4557.2010.00360.x

1227 Miller-Fleming L, Giorgini F, Outeiro TF (2008) Yeast as a model for studying human neurodegenerative  
 1228 disorders. Biotechnol J 3:325–338. doi: 10.1002/biot.200700217

1229 Milreu PV, Klein CC, Cottret L, et al (2014) Telling metabolic stories to explore metabolomics data: a case  
 1230 study on the yeast response to cadmium exposure. Bioinformatics 30:61–70. doi:  
 1231 10.1093/bioinformatics/btt597

1232 Morgera E, Tsioumani E, Buck M (2014) Unraveling the Nagoya Protocol. Brill, Leiden, The Netherlands

1233 Moulton GG (2014) Fed-Batch Fermentation. Woodhead Publishing, Cambridge, UK

1234 Moyer RA, Hummer KE, Finn CE, et al (2001) Anthocyanins, Phenolics, and Antioxidant Capacity in  
 1235 Diverse Small Fruits: *Vaccinium*, *Rubus*, and *Ribes*. J Agric Food Chem 50:519–525. doi:  
 1236 10.1021/JF011062R

1237 Määttä KR, Kamal-Eldin A, Törrönen AR (2003) High-Performance Liquid Chromatography (HPLC)  
 1238 Analysis of Phenolic Compounds in Berries with Diode Array and Electrospray Ionization Mass  
 1239 Spectrometric (MS) Detection: *Ribes* Species. J Agric Food Chem 51:6736–6744. doi:  
 1240 10.1021/JF0347517

1241 Määttä KR, Kamal-Eldin A, Törrönen R (2001) Phenolic Compounds in Berries of Black, Red, Green, and  
 1242 White Currants ( *Ribes* sp.). Antioxid Redox Signal 3:981–993. doi: 10.1089/152308601317203521

1243 Neves AR, Pool WA, Kok J, et al (2005) Overview on sugar metabolism and its control in *Lactococcus lactis*  
 1244 — The input from in vivo NMR. FEMS Microbiol Rev 29:531–554. doi:  
 1245 <http://dx.doi.org/10.1016/j.fmrre.2005.04.005>

1246 NMRO (UK) (2015) The Nagoya protocol on access and benefit sharing of genetic resources: compliance  
 1247 and guidance.

- 1248 O'Brien RJ, Wong PC (2011) Amyloid precursor protein processing and Alzheimer's disease. *Annu Rev*  
1249 *Neurosci* 34:185–204. doi: 10.1146/annurev-neuro-061010-113613
- 1250 Oliveira AP, Nielsen J, Förster J (2005) Modeling *Lactococcus lactis* using a genome-scale flux model.  
1251 *BMC Microbiol* 5:39. doi: 10.1186/1471-2180-5-39
- 1252 Outeiro TF, Lindquist S (2003) Yeast cells provide insight into alpha-synuclein biology and pathobiology.  
1253 *Science* 302:1772–5. doi: 10.1126/science.1090439
- 1254 Palmieri N, Forleo MB, Salimei E (2017) Environmental impacts of a dairy cheese chain including whey  
1255 feeding: An Italian case study. *J Clean Prod* 140:881–889. doi: 10.1016/j.jclepro.2016.06.185
- 1256 Pan M-G, Xiong Y, Chen F (2013) NFAT gene family in inflammation and cancer. *Curr Mol Med* 13:543–  
1257 54.
- 1258 Pandey RP, Parajuli P, Koffas MAG, Sohng JK (2016) Microbial production of natural and non-natural  
1259 flavonoids: Pathway engineering, directed evolution and systems/synthetic biology. *Biotechnol Adv*  
1260 34:634–662. doi: 10.1016/j.biotechadv.2016.02.012
- 1261 Paredes-López O, Cervantes-Ceja ML, Vigna-Pérez M, Hernández-Pérez TT (2010) Berries: improving  
1262 human health and healthy aging, and promoting quality life - a review. *Plant Foods Hum Nutr* 65:299–  
1263 308. doi: 10.1007/s11130-010-0177-1
- 1264 Pátek M, Nešvera J (2013) Promoters and Plasmid Vectors of *Corynebacterium glutamicum*. In: Yukawa H,  
1265 Inui M (eds) *Corynebacterium glutamicum: Biology and Biotechnology*, 23rd edn. Springer, Berlin,  
1266 Germany, pp 51–88
- 1267 Patil K, Rocha I, Förster J, et al (2005) Evolutionary programming as a platform for in silico metabolic  
1268 engineering. *BMC Bioinformatics* 6:308. doi: 10.1186/1471-2105-6-308
- 1269 Pauli GF, Chen S-N, Friesen JB, et al (2012) Analysis and Purification of Bioactive Natural Products: The

1270 AnaPurNa Study. J Nat Prod 75:1243–1255. doi: 10.1021/np300066q

1271 Pedersen MB, Gaudu P, Lechardeur D, et al (2012) Aerobic Respiration Metabolism in Lactic Acid Bacteria  
 1272 and Uses in Biotechnology. Annu Rev Food Sci Technol 3:37–58. doi: 10.1146/annurev-food-022811-  
 1273 101255

1274 Pei L, Schmidt M (2016) Sustainable Assessment on Using Bacterial Platform to Produce High-Added-  
 1275 Value Products from Berries through Metabolic Engineering. In: Singh S (ed) Systems Biology  
 1276 Application in Synthetic Biology. Springer India, pp 71–78

1277 Pontes DS, de Azevedo MSP, Chatel J-M, et al (2011) *Lactococcus lactis* as a live vector: Heterologous  
 1278 protein production and DNA delivery systems. Protein Expr Purif 79:165–175. doi:  
 1279 10.1016/j.pep.2011.06.005

1280 Prescott TAK, Ariño J, Kite GC, Simmonds MSJ (2012) Inhibition of human calcineurin and yeast  
 1281 calcineurin-dependent gene expression by *Jasminum humile* leaf and root extracts. J Ethnopharmacol  
 1282 140:293–297. doi: 10.1016/j.jep.2012.01.020

1283 Prince K, Smith M (2004) Protein Purification Protocols. In: Cutler P (ed) Protein Purification Protocols.  
 1284 Humana Press, Totowa, USA, pp 463–480

1285 Quideau S, Deffieux D, Douat-Casassus C, Pouységu L (2011) Plant Polyphenols: Chemical Properties,  
 1286 Biological Activities, and Synthesis. Angew Chemie Int Ed 50:586–621. doi: 10.1002/anie.201000044

1287 Rani MU, Appaiah A (2011) Optimization of culture conditions for bacterial cellulose production from  
 1288 *Gluconacetobacter hansenii* UAC09. Ann Microbiol 61:781–787. doi: 10.1007/s13213-011-0196-7

1289 Rasmussen SE, Frederiksen H, Struntze Krogholm K, Poulsen L (2005) Dietary proanthocyanidins:  
 1290 Occurrence, dietary intake, bioavailability, and protection against cardiovascular disease. Mol Nutr  
 1291 Food Res 49:159–174. doi: 10.1002/mnfr.200400082

- 1292 Ratnam BVV, Narasimha Rao M, Damodar Rao M, et al (2003) Optimization of fermentation conditions for  
1293 the production of ethanol from sago starch using response surface methodology. *World J Microbiol*  
1294 *Biotechnol* 19:523–526. doi: 10.1023/A:1025174731814
- 1295 Renaud S, de Lorgeril M (1992) Wine, alcohol, platelets, and the French paradox for coronary heart disease.  
1296 *Lancet* 339:1523–1526. doi: 10.1016/0140-6736(92)91277-F
- 1297 Ro D-K, Paradise EM, Ouellet M, et al (2006) Production of the antimalarial drug precursor artemisinic acid  
1298 in engineered yeast. *Nature* 440:940–943. doi: 10.1038/nature04640
- 1299 Ruiz A, Hermosín-Gutiérrez I, Mardones C, et al (2010) Polyphenols and Antioxidant Activity of Calafate  
1300 (*Berberis microphylla*) Fruits and Other Native Berries from Southern Chile. *J Agric Food Chem*  
1301 58:6081–6089. doi: 10.1021/jf100173x
- 1302 Rusnak F, Mertz P (2000) Calcineurin: form and function. *Physiol Rev* 80:1483–521.
- 1303 Santos CNS, Koffas M, Stephanopoulos G (2011) Optimization of a heterologous pathway for the  
1304 production of flavonoids from glucose. *Metab Eng* 13:392–400. doi: 10.1016/j.ymben.2011.02.002
- 1305 SCBD (2011) Nagoya Protocol on access to genetic resources and the fair and equitable sharing of benefits  
1306 arising from their utilization to the convention on biological diversity. Quebec, Canada
- 1307 Secretary of State (UK) (2015) Nagoya Protocol (Compliance) Regulations 2015.
- 1308 Segrè D, Vitkup D, Church GM (2002) Analysis of optimality in natural and perturbed metabolic networks.  
1309 *Proc Natl Acad Sci U S A* 99:15112–7. doi: 10.1073/pnas.232349399
- 1310 Shen X-H, Zhou N-Y, Liu S-J (2012) Degradation and assimilation of aromatic compounds by  
1311 *Corynebacterium glutamicum*: another potential for applications for this bacterium? *Appl Microbiol*  
1312 *Biotechnol* 95:77–89. doi: 10.1007/s00253-012-4139-4
- 1313 Shen X, Liu S (2005) Key enzymes of the protocatechuate branch of the  $\beta$ -ketoadipate pathway for aromatic



1314 degradation in *Corynebacterium glutamicum*. Sci China Ser C Life Sci 48:241–249. doi:  
 1315 10.1007/BF03183617

1316 Shinfuku Y, Sorpitiporn N, Sono M, et al (2009) Development and experimental verification of a genome-  
 1317 scale metabolic model for *Corynebacterium glutamicum*. Microb Cell Fact 8:43. doi: 10.1186/1475-  
 1318 2859-8-43

1319 Shults CW (2006) Lewy bodies. Proc Natl Acad Sci U S A 103:1661–8. doi: 10.1073/pnas.0509567103

1320 Simeonidis E, Price ND (2015) Genome-scale modeling for metabolic engineering. J Ind Microbiol  
 1321 Biotechnol 42:327–338. doi: 10.1007/s10295-014-1576-3

1322 Song AA, Abdullah JO, Abdullah MP, et al (2014) Engineering the lactococcal mevalonate pathway for  
 1323 increased sesquiterpene production. FEMS Microbiol Lett 355:177–84. doi: 10.1111/1574-6968.12469

1324 Stahlhut SG, Siedler S, Malla S, et al (2015) Assembly of a novel biosynthetic pathway for production of the  
 1325 plant flavonoid fisetin in *Escherichia coli*. Metab Eng 31:84–93. doi: 10.1016/j.ymben.2015.07.002

1326 Stanford NJ, Millard P, Swainston N (2015) RobOKoD: microbial strain design for (over)production of  
 1327 target compounds. Front Cell Dev Biol 3:17. doi: 10.3389/fcell.2015.00017

1328 Stavang JA, Freitag S, Foito A, et al (2015) Raspberry fruit quality changes during ripening and storage as  
 1329 assessed by colour, sensory evaluation and chemical analyses. Sci Hortic (Amsterdam) 195:216–225.  
 1330 doi: 10.1016/j.scienta.2015.08.045

1331 Stracke R, Werber M, Weisshaar B (2001) The R2R3-MYB gene family in *Arabidopsis thaliana*. Curr Opin  
 1332 Plant Biol 4:447–456. doi: 10.1016/S1369-5266(00)00199-0

1333 Su LJ, Auluck PK, Outeiro TF, et al (2010) Compounds from an unbiased chemical screen reverse both ER-  
 1334 to-Golgi trafficking defects and mitochondrial dysfunction in Parkinson’s disease models. Dis Model  
 1335 Mech 3:194–208. doi: 10.1242/dmm.004267

- 1336 Suh Y, Afaq F, Johnson JJ, Mukhtar H (2008) A plant flavonoid fisetin induces apoptosis in colon cancer  
1337 cells by inhibition of COX2 and Wnt/EGFR/NF- B-signaling pathways. *Carcinogenesis* 30:300–307.  
1338 doi: 10.1093/carcin/bgn269
- 1339 Tardiff DF, Tucci ML, Caldwell KA, et al (2012) Different 8-hydroxyquinolines protect models of TDP-43  
1340 protein,  $\alpha$ -synuclein, and polyglutamine proteotoxicity through distinct mechanisms. *J Biol Chem*  
1341 287:4107–20. doi: 10.1074/jbc.M111.308668
- 1342 Tayarani-Najaran Z, Mousavi SH, Tajfard F, et al (2013) Cytotoxic and apoptogenic properties of three  
1343 isolated diterpenoids from *Salvia chorassanica* through bioassay-guided fractionation. *Food Chem*  
1344 *Toxicol* 57:346–351. doi: 10.1016/j.fct.2013.03.037
- 1345 Tepper N, Shlomi T (2010) Predicting metabolic engineering knockout strategies for chemical production:  
1346 accounting for competing pathways. *Bioinformatics* 26:536–43. doi: 10.1093/bioinformatics/btp704
- 1347 Tohge T, Nishiyama Y, Hirai MY, et al (2005) Functional genomics by integrated analysis of metabolome  
1348 and transcriptome of *Arabidopsis* plants over-expressing an MYB transcription factor. *Plant J* 42:218–  
1349 235. doi: 10.1111/j.1365-313X.2005.02371.x
- 1350 Törrönen R (2009) Sources and health effects of dietary ellagitannins. In: Quideau S (ed) *Chemistry and*  
1351 *Biology of Ellagitannins: An Underestimated Class of Bioactive Plant Polyphenols*. World Scientific  
1352 Publishing, Singapore, p 298–319
- 1353 Vagiri M, Ekholm A, Öberg E, et al (2013) Phenols and ascorbic acid in black currants (*Ribes nigrum* L.):  
1354 variation due to genotype, location, and year. *J Agric Food Chem* 61:9298–306. doi: 10.1021/jf402891s
- 1355 van Summeren-Wesenhagen P V, Marienhagen J (2015) Metabolic Engineering of *Escherichia coli* for the  
1356 Synthesis of the Plant Polyphenol Pinosylvin. *Appl Environ Microbiol* 81:840–849. doi:  
1357 10.1128/AEM.02966-14
- 1358 Vance C, Rogelj B, Hortobágyi T, et al (2009) Mutations in FUS, an RNA processing protein, cause familial

1359 amyotrophic lateral sclerosis type 6. *Science* 323:1208–11. doi: 10.1126/science.1165942

1360 Vazquez-Albacete D, Cavaleiro AM, Christensen U, et al (2016) An expression tag toolbox for microbial  
 1361 production of membrane bound plant cytochromes P450. *Biotechnol Bioeng*. doi: 10.1002/bit.26203

1362 Vera J, González-Alcón C, Marín-Sanguino A, Torres N (2010) Optimization of biochemical systems  
 1363 through mathematical programming: Methods and applications. *Comput Oper Res* 37:1427–1438. doi:  
 1364 10.1016/j.cor.2009.02.021

1365 Veríssimo A, Bassard J-E, Julien-Laferrrière A, et al MassBlast: A workflow to accelerate RNA-seq and  
 1366 DNA database analysis.

1367 Vogt T (2010) Phenylpropanoid Biosynthesis. *Mol Plant* 3:2–20. doi: 10.1093/mp/ssp106

1368 Waldo GS, Standish BM, Berendzen J, Terwilliger TC (1999) Rapid protein-folding assay using green  
 1369 fluorescent protein. *Nat Biotechnol* 17:691–695. doi: 10.1038/10904

1370 Wang J, Guleria S, Koffas MA, Yan Y (2016a) Microbial production of value-added nutraceuticals. *Curr*  
 1371 *Opin Biotechnol* 37:97–104. doi: 10.1016/j.copbio.2015.11.003

1372 Wang Y, Bhuiya MW, Zhou R, Yu O (2015) Pterostilbene production by microorganisms expressing  
 1373 resveratrol O-methyltransferase. *Ann Microbiol* 65:817–826. doi: 10.1007/s13213-014-0922-z

1374 Wang Y, Zhu J, Meng X, et al (2016b) Comparison of polyphenol, anthocyanin and antioxidant capacity in  
 1375 four varieties of *Lonicera caerulea* berry extracts. *Food Chem* 197:522–529. doi:  
 1376 10.1016/j.foodchem.2015.11.006

1377 Watts KR, Ratnam J, Ang K-H, et al (2010) Assessing the trypanocidal potential of natural and semi-  
 1378 synthetic diketopiperazines from two deep water marine-derived fungi. *Bioorg Med Chem* 18:2566–  
 1379 2574. doi: 10.1016/j.bmc.2010.02.034

1380 Weckwerth W (2003) Metabolomics in systems biology. *Annu Rev Plant Biol* 54:669–689. doi:

10.1146/annurev.arplant.54.031902.135014

Weller MG, G. M (2012) A Unifying Review of Bioassay-Guided Fractionation, Effect-Directed Analysis and Related Techniques. *Sensors* 12:9181–9209. doi: 10.3390/s120709181

Wen Z, Minton NP, Zhang Y, et al (2016) Enhanced solvent production by metabolic engineering of a twin-clostridial consortium. *Metab Eng*. doi: 10.1016/j.ymben.2016.10.013

Xu P, Ranganathan S, Fowler ZL, et al (2011) Genome-scale metabolic network modeling results in minimal interventions that cooperatively force carbon flux towards malonyl-CoA. *Metab Eng* 13:578–587. doi: 10.1016/j.ymben.2011.06.008

Xu P, Wang W, Li L, et al (2013) Design and Kinetic Analysis of a Hybrid Promoter–Regulator System for Malonyl-CoA Sensing in *Escherichia coli*. doi: 10.1021/cb400623m

Yang X, Summerhurst DK, Koval SF, et al (2001) Isolation of an antimicrobial compound from *Impatiens balsamina* L. using bioassay-guided fractionation. *Phyther Res* 15:676–680. doi: 10.1002/ptr.906

Yu C, Cao Y, Zou H, Xian M (2011) Metabolic engineering of *Escherichia coli* for biotechnological production of high-value organic acids and alcohols. *Appl Microbiol Biotechnol* 89:573–583. doi: 10.1007/s00253-010-2970-z

Zadernowski R, Naczki M, Nesterowicz J (2005) Phenolic Acid Profiles in Some Small Berries. *J Agric Food Chem* 53:2118–2124. doi: 10.1021/JF040411P

Zelle RM, de Hulster E, Kloezen W, et al (2010) Key process conditions for production of C(4) dicarboxylic acids in bioreactor batch cultures of an engineered *Saccharomyces cerevisiae* strain. *Appl Environ Microbiol* 76:744–50. doi: 10.1128/AEM.02396-09

Zhang H, Matsuda H, Yamashita C, et al (2009) Hydrangeic acid from the processed leaves of *Hydrangea macrophylla* var. *thunbergii* as a new type of anti-diabetic compound. *Eur J Pharmacol* 606:255–261.

1403           doi: 10.1016/j.ejphar.2009.01.005

1404   Zheng J, Ding C, Wang L, et al (2011) Anthocyanins composition and antioxidant activity of wild *Lycium*

1405           *rutenicum* Murr. from Qinghai-Tibet Plateau. Food Chem 126:859–865. doi:

1406           10.1016/j.foodchem.2010.11.052

1407   Zomorodi AR, Suthers PF, Ranganathan S, Maranas CD (2012) Mathematical optimization applications in

1408           metabolic networks. Metab Eng 14:672–686. doi: 10.1016/j.ymben.2012.09.005

1409

1410

1411 **Figure legends**

1412 **Figure 1:** **a)** The molecular structure of the stilbene *trans*-resveratrol. **b)** Basic flavonoid structure. R<sub>x</sub> are  
1413 positions where decorating groups (*e.g.* hydroxyl, acyl, glycosyl, etc.), which differentiate the compounds  
1414 and alter their chemical properties, could be attached.

1415 **Figure 2:** Principal component analysis (PCA) plots of the full dataset without reference material and  
1416 blanks. Each dot represents the profile of one sample aliquot whereas lines delimit groups of samples that  
1417 cluster together. **a)** PCA plots of the full dataset without reference material and blanks in positive mode.  
1418 Components (Scr) 1 and 2 explain 9.79% and 7.59% of the variation respectively. **b)** PCA plots of the full  
1419 dataset without reference material and blanks in negative mode. Components (Scr) 1 and 2 explain 13.20%  
1420 and 10.30% of the variation respectively.

1421 **Figure 3:** Schematic representation of yeast models. **a.** Yeast models of neurodegenerative diseases.  
1422 Humanized yeast strains encoding chimeric fusions of disease genes with GFP (green fluorescent protein)  
1423 under the control of a galactose-inducible (GAL1) promoter. Expression of A $\beta$ 42,  $\alpha$ Syn (SNCA gene), FUS  
1424 and Htt<sup>pQ103</sup> leads to the formation of toxic aggregates, recapitulating the pathological processes of  
1425 Alzheimer's disease (AD), Parkinson's disease (PD), Amyotrophic lateral sclerosis (ALS), and Huntington's  
1426 disease (HD), respectively. **b.** Yeast model of Crz1 (NFAT orthologue) activation. Crz1 is activated by the  
1427 serine/threonine protein phosphatase calcineurin (composed by CnA and CnB subunits). The reporter strain  
1428 encodes the *lacZ* gene under the control of a promoter containing Crz1-binding sites, allowing the  
1429 assessment of Crz1 activation through the measurement of  $\beta$ -galactosidase activity.

1430 **Figure 4:** Identification of bioactivities in the berry germplasm. **a)** Strategy used for the discovery of  
1431 bioactive compounds for Alzheimer's disease. BY4741\_erg6 recombinant yeast cells expressing GFP-A $\beta$ 42  
1432 were grown in SD galactose medium and cell viability was assessed by growth curves for 24 h in cells  
1433 challenged or unchallenged with *Lycium chinense* polyphenol extracts. Cells containing the empty vector  
1434 were used as controls. The areas under the curve (AUC) were integrated using the Origin software

1435 (OriginLab, Northampton, MA). Values represent the mean  $\pm$  standard error of mean (SEM) of at least three  
1436 biological replicates.

1437 b) Strategy used for the discovery of bioactive compounds with anti-inflammatory potential. BY4741 yeast  
1438 cells encoding CDRE<sub>promoter</sub>-*lacZ* were treated or not with *Lycium chinense* polyphenol extracts and induced  
1439 with 1.8 mM MnCl<sub>2</sub>. Crz1 activation was assessed by monitoring  $\beta$ -galactosidase activity using ortho-  
1440 nitrophenyl- $\beta$ -galactoside (ONPG). The immunosuppressant FK506, a well-known inhibitor of calcineurin  
1441 (Dumont 2000) was used as a positive control. The values represent the mean  $\pm$  SEM of at least three  
1442 biological replicates, \**p* < 0.05, \*\**p* < 0.01, \*\*\**p* < 0.001.

1443 **Figure 5.** Yeast as a discovery platform of therapeutic compounds. Iterative bio-guided fractionation of  
1444 complex mixtures allows the discovery of small molecules improving cell growth, by rescue of protein  
1445 aggregates toxicity, or inhibiting inflammatory processes. High-throughput formats aligned with chemical  
1446 identification by LC-MS approaches accelerate the identification of candidate compounds.

1447 **Figure 6:** Polyphenol biosynthetic pathway in plants (modified from Falcone Ferreyra et al. 2012). PAL,  
1448 phenylalanine ammonia-lyase; C4H, cinnamate 4-hydroxylase; 4CL, 4-coumaroyl-CoA ligase; CHS,  
1449 chalcone synthase; CHI, chalcone isomerase; F3H, flavanone 3-hydroxylase; F3'5'H, flavonoid 3'-5'-  
1450 hydroxylase; DFR, dihydroflavonol 4-reductase; ANS, anthocyanidin synthase; ANR, anthocyanidin  
1451 reductase; UFGT, flavonoid 3-*O*-glucosyltransferase. LAR, leucoanthocyanidin reductase; FNS, flavone  
1452 synthase; FLS, flavonol synthase; STS, stilbene synthase. P450 enzymes are highlighted in red.

1453 **Figure 7:** Overview of the endogenous phenylpropanoid degradation and the engineered pathway for  
1454 polyphenol synthesis in *C. glutamicum*. One of the precursors for the production of the stilbene *trans*-  
1455 resveratrol (catalyzed by stilbene synthase, STS) or for naringenin chalcone (catalyzed by chalcone  
1456 synthases, CHS) is *p*-coumaroyl-CoA. Naringenin chalcone isomerizes to the (2*S*)-flavanone naringenin  
1457 either spontaneously or catalyzed by the activity of chalcone isomerase (CHI). In *C. glutamicum*, *p*-coumaric  
1458 acid can be degraded to 4-hydroxybenzoate by a CoA-dependent,  $\beta$ -oxidative deacetylation pathway. 4-  
1459 Hydroxybenzoate is subsequently hydroxylated to protocatechuate, which is catabolized to succinyl-CoA

1460 and acetyl-CoA by the  $\beta$ -ketoadipate pathway. Genes coding for the underlined enzymes were deleted in the  
1461 course of the construction of the platform strain *C. glutamicum* DelAro<sup>4</sup>.

1462 **Figure 8:** Heterologous biosynthetic pathway for fisetin assembled in *E. coli*. TAL, tyrosine ammonia-lyase;  
1463 4CL, 4-coumaroyl-CoA ligase; CHS::CHR, chalcone synthase::chalcone reductase fusion; CHI, chalcone  
1464 isomerase; F3H, flavanone 3-hydroxylase; FLS, flavonol synthase; F3'H::CPR, flavonoid 3'-  
1465 hydroxylase::cytochrome P450 reductase fusion.

1466 **Figure 9:** Heterologous biosynthetic pathways introduced into the genome-scale models of *C. glutamicum*  
1467 and *L. lactis* leading to the four target compounds: i) *trans*-resveratrol ii) pelargonidin iii) quercetin and iv)  
1468 fisetin. Two heterologous enzymes (TAL and 4CL) are common for the biosynthesis of the phenolic  
1469 molecules. Enzyme abbreviations: TAL, tyrosine ammonia-lyase; 4CL, 4-coumaroyl-CoA ligase; STS,  
1470 stilbene synthase; CHS, chalcone synthase; CHR, chalcone reductase; CHI, chalcone isomerase; F3H,  
1471 flavanone 3-hydroxylase; FLS, flavonol synthase; F3'H, flavonoid 3' hydroxylase; CPR, cytochrome P450  
1472 reductase.

1473 **Figure 10:** Partition coefficients in the form of log P values of *trans*-resveratrol using different organic  
1474 solvents using *C. glutamicum* fermentation broth as the aqueous phase.

1475 **Figure 11:** On the left, protein concentration (expressed as percentage of the initial broth protein  
1476 concentration) in the organic and aqueous phases, after performing liquid extraction. The concentration in  
1477 the organic phase is almost zero (considering the error bars), indicating that some protein stayed in the  
1478 interphase. On the right, the formation of an interphase after the broth had been extracted with organic  
1479 solvent (a mixture of heptanoic acid and hexyl acetate in this case) was clear.

1480 **Figure 12:** Chromatogram following a pulse injection of 10 ml of *C. glutamicum* spent fermentation broth in  
1481 a 30 ml column packed with Amberlite XAD-16 resin. A washing step was performed with 10 column  
1482 volumes (CVs) of MilliQ water. Afterwards, elution was performed with a water/ethanol gradient for 12 CV,  
1483 followed by a final regeneration step with 100% ethanol.

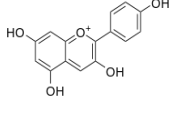
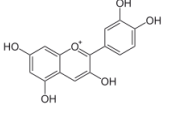
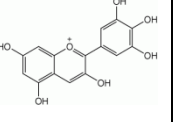
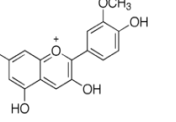
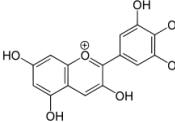
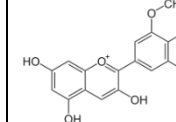


1484 **Figure 13:** Experimental set-up for the “Continuous Product Removal” mode.

1485

1486      **Tables**

1487      **Table 1:** Anthocyanidin structures (aglycone form) found in different berry species. The presence of a  
1488      compound is noted with + whereas the absence is noted with -.

Type	Pelargonidin	Cyanidin	Delphinidin	Peonidin	Petunidin	Malvinidin
Structure						
Molecular formula	C <sub>15</sub> H <sub>11</sub> O <sub>5</sub> <sup>+</sup>	C <sub>15</sub> H <sub>11</sub> O <sub>6</sub> <sup>+</sup>	C <sub>15</sub> H <sub>11</sub> O <sub>7</sub> <sup>+</sup>	C <sub>16</sub> H <sub>13</sub> O <sub>6</sub> <sup>+</sup>	C <sub>16</sub> H <sub>13</sub> O <sub>7</sub> <sup>+</sup>	C <sub>17</sub> H <sub>15</sub> O <sub>7</sub> <sup>+</sup>
<i>Fragaria</i> <sup>1</sup>	+	+	-	-	-	-
<i>Ribes</i> <sup>1</sup>	+	+	+	-	+	-
<i>Rubus</i> <sup>1</sup>	+	+	-	-	-	-
<i>Vaccinium</i> <sup>1</sup>	-	+	+	+	+	+
<i>Vitis</i> <sup>1</sup>	-	+	+	+	+	+
<i>Lonicera</i> <sup>2,5</sup>	+	+	+	+	-	+
<i>Lycium</i> <sup>3</sup>	-	-	+	-	+	+
<i>Berberis</i> <sup>4</sup>	-	+	+	+	+	+
<i>Ugni</i> <sup>4</sup>	-	+	-	+	-	-
<i>Aristotelia</i> <sup>6</sup>	-	+	+	-	-	-

1489      <sup>1</sup> Phenol explorer database (<http://phenol-explorer.eu>); <sup>2</sup> (Chaovanalikit et al. 2004); <sup>3</sup> (Zheng et al. 2011); <sup>4</sup>  
1490      (Ruiz et al. 2010); <sup>5</sup> (Wang et al. 2016b); <sup>6</sup> (Fredes et al. 2014)

1491

1492

1493 **Table 2:** Selected species for bioactivity testing according to the phytochemical diversity.

Species	Variety/line	Species	Variety/line
<i>Aristotelia chilensis</i> (Molina)Stuntz	-	<i>Rubus armeniacus</i> Focke	Himalayan Giant
<i>Berberis buxifolia</i> Lam.	-	<i>Rubus occidentalis</i> L.	Huron
<i>Lonicera caerulea</i> L.	S322-23	<i>Rubus loganobaccus</i> L.H.Bailey	Sunberry
<i>Lycium chinense</i> Mill.	-	<i>Rubus loganobaccus</i> L.H.Bailey	Tayberry
<i>Ribes grossularia</i> L.	9294	<i>Rubus bartonii</i> Newton	Ashton Cross
<i>Ribes grossularoides</i> Maxim.	H1-12-1	<i>Rubus fruticosus</i> L. agg.	Fantasia
<i>Ribes nevadense</i> Kellog	-	<i>Rubus brigantinus</i> Samp.	-
<i>Ribes spp.</i>	644217	<i>Rubus genevierii</i> Boreau.	-
<i>Ribes spp.</i>	1126	<i>Rubus henriquesii</i> Samp.	-
<i>Ribes spp.</i>	2292-1	<i>Rubus hochstetterorum</i> Seub.	-
<i>Ribes spp.</i>	Muravushka	<i>Rubus sampaioanus</i> Sudre ex Samp.	-
<i>Ribes spp.</i>	-	<i>Rubus vagabundus</i> Samp.	-
<i>Ribes petraeum</i> Wulf.	R2-1-2	<i>Rubus idaeus</i> L.	Octavia
<i>Ribes x saundersii</i> Jancz.	XB4	<i>Rubus idaeus</i> L.	Prestige
<i>Ribes bethmontii</i> Jancz.	XW6	<i>Ugni molinae</i> Turcz.	Red Pearl
<i>Ribes rubrum</i> L.	S11-3-36	<i>Vaccinium uliginosum</i> L.	-
<i>Rubus spp.</i>	B14	<i>Vaccinium vitis-idaea</i> L.	-

1494

1495

1496 **Table 3:** Transcriptome resources for berries developed on BacHBerry

Species	Plant material	Total number of raw reads
<i>Aristotelia sp.</i>	ripe berries	397,707,372
<i>Berberis sp.</i>	ripe berries	444,362,698
<i>Corema sp.</i>	leaf	353,604,932
<i>Lonicera sp.</i>	ripe berries	397,214,254
<i>Ribes sp.1</i>	ripe berries	336,479,242
<i>Ribes sp.2</i>	ripe berries	393,665,630
<i>Rubus sp.1</i>	berries (3 stages)	1,040,224,680
<i>Rubus sp.2</i>	berries (3 stages)	1,064,858,518
<i>Rubus sp.3</i>	ripe berries	505,754,030
<i>Rubus sp.4</i>	ripe berries	390,608,452
<i>Ugni sp.</i>	ripe berries	405,024,920
<i>Vaccinium sp.1</i>	ripe berries	373,159,882
<i>Vaccinium sp.2</i>	ripe berries	375,778,718

1497

1498

1499 **Table 4:** Statistics on five major enzyme families representing conserved biochemical steps of the  
 1500 phenylpropanoid pathway. Candidates displayed more than 70% sequence coverage to reference query  
 1501 sequences. This work led to the creation of a large dataset of candidates used for pathway elucidation and  
 1502 reconstitution in bacterial hosts.

Enzyme families	Candidates per transcriptome (min-max)
Conserved phenylpropanoid pathway enzymes	109-180
CYPs	100-177
UGTs	70-145
2-OGDs	32-88
<i>O</i> -methyltransferases	8-51
Acyltransferases	23-36

1503

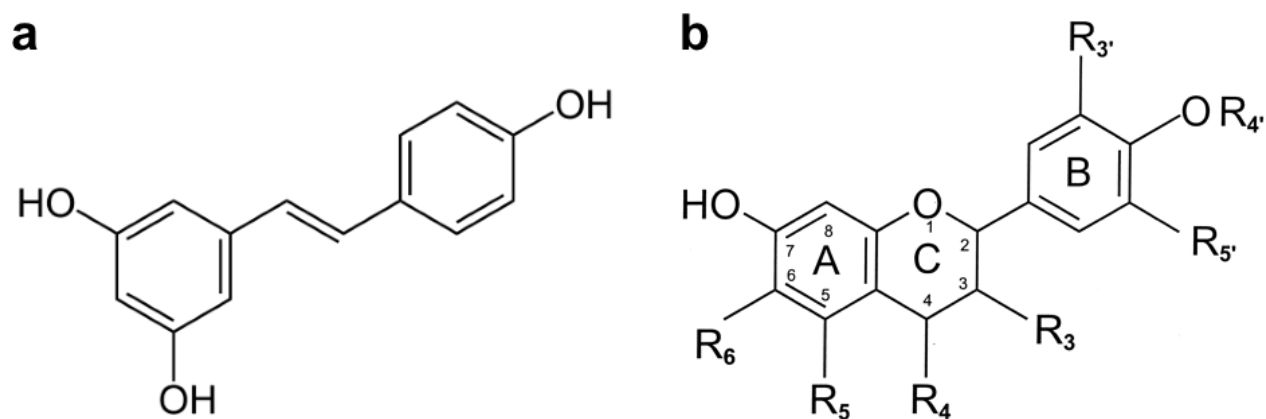
1504

Specific application area in project	Bioprospecting for discovery of novel berry phenolics	Characterization of phenolics biosynthetic pathways	Bacterial cell factories for production of berry phenolics	Fermentation and bioprocess engineering
General aims of BacHBerry	To tap into the biodiversity of berries from around the globe and of their phenolics content	Identification and characterization of berry phenolics biosynthetic pathways	Design and generation of bacterial cell factories for production of high-value berry phenolics	Implementation of production of high-value phenolics in fermenters up to demonstration scale
Achievements of the project	<ul style="list-style-type: none"> <li>• Standardization of methodologies for harvesting, extract preparation, phenolics content assessment and fractionation</li> <li>• Metabolomics data from berry species originating from UK, mainland Europe, Russia, Chile, and China was obtained and made available in form of an online database</li> <li>• Implementation of robust assays based on the SMART platform for discovery of bioactivity and functionality</li> </ul>	<ul style="list-style-type: none"> <li>• Transcriptomes of 13 germplasms spanning eight genera, seven families and seven orders were generated and analyzed</li> <li>• New algorithms for functional genomics and improved computational methods for pathway identification were established</li> <li>• Over 4000 candidate genes for various biosynthetic steps, transport and regulation of polyphenol production were identified</li> </ul>	<ul style="list-style-type: none"> <li>• <i>L. lactis</i> strains were engineered for stilbene (<i>trans</i>-resveratrol and pterostilbene) and anthocyanin production</li> <li>• Production of various stilbenes (<i>trans</i>-resveratrol, pinosylvin, and piceatannol), methylated stilbenes (pinostilbene and pterostilbene), flavanones (naringenin, pinocembrin, and eryodictiol), and flavonols (kaempferol and quercetin) was achieved in <i>C. glutamicum</i> with yields comparable to those obtained in the model organism <i>E. coli</i>. Production titers for the produced</li> </ul>	<ul style="list-style-type: none"> <li>• Cultivation conditions for improved productivity were developed for both <i>L. lactis</i> and <i>C. glutamicum</i> through optimization of fermentation parameters (batch versus fed-batch operation, substrate and dissolved oxygen concentration, etc.)</li> <li>• Whey was evaluated as an alternative carbon source for polyphenol production in <i>L. lactis</i></li> <li>• Multiple bioseparation methods for extraction of produced polyphenols from fermentation broth were evaluated</li> </ul>

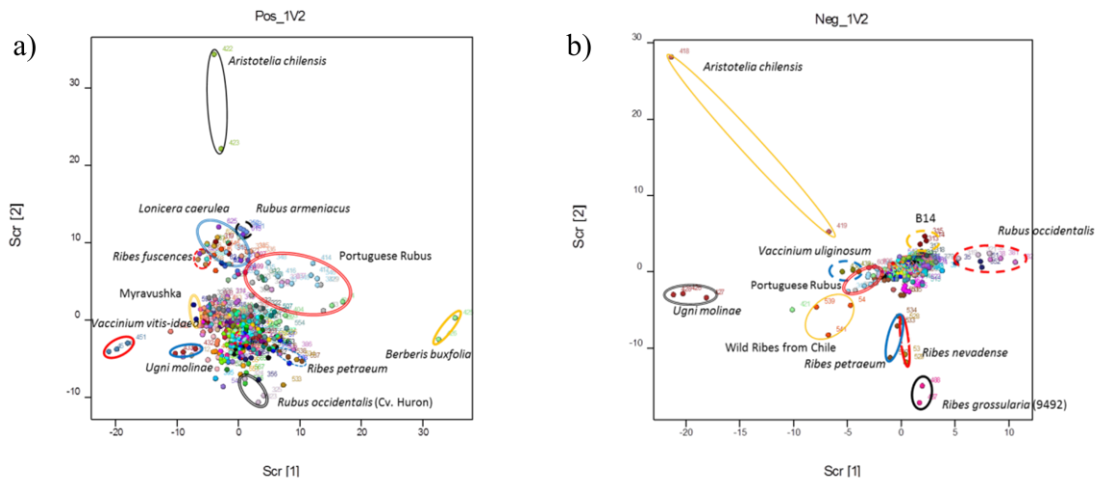
	<ul style="list-style-type: none"> <li>• Obtained berry extracts were assayed for bioactivity against several human diseases, and additionally evaluated for other functional uses (eq. food additives, antimicrobials). Bioactivities against Alzheimer's, Parkinson's and Huntington's diseases, Amyotrophic Lateral Sclerosis and inflammation were detected.</li> <li>• Several pure bioactive effector compounds were identified and validated</li> </ul>	<ul style="list-style-type: none"> <li>• Biochemical activities of some of the gene candidates were tested, thus validating the predictions</li> </ul>	<p>flavonols are the highest titers obtained in engineered microorganism until today</p> <ul style="list-style-type: none"> <li>• Polyphenol production was further improved by engineering enhanced precursor supply (eq. L-tyrosine, malonyl-CoA) via rational design, modeling-based prediction, and use of biosensors</li> <li>• Alternative biosynthetic routes were explored though engineering of <i>trans</i>-resveratrol production from 4-hydroxybenzoic acid (HBA) in <i>C. glutamicum</i></li> <li>• Computational models were developed allowing predicting a minimal set of modifications needed for improved polyphenol production in both a single-strain population and a consortium of two or more species</li> </ul>	<ul style="list-style-type: none"> <li>• Two possible designs for a system for continuous product removal from the fermentation broth were implemented</li> <li>• Resveratrol production in <i>C. glutamicum</i> was scaled up from shake flask to 5 l fermenter to demonstration scale (250 l)</li> </ul>
--	--	--	---	--





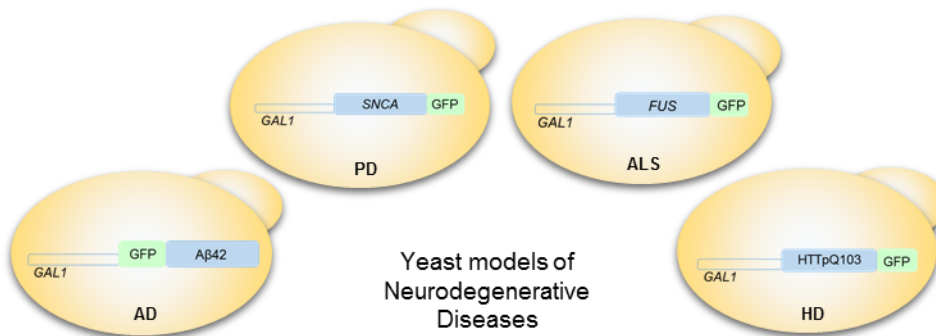


**Figure 1:** **a)** The molecular structure of the stilbene *trans*-resveratrol. **b)** Basic flavonoid structure. R<sub>x</sub> are positions where decorating groups (*e.g.* hydroxyl, acyl, glycosyl, etc.), which differentiate the compounds and alter their chemical properties, could be attached.

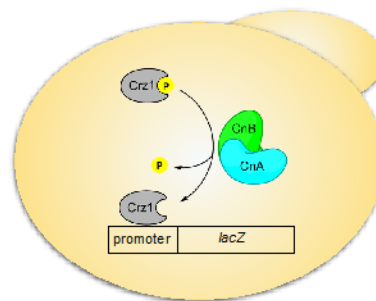


**Figure 2:** Principal component analysis (PCA) plots of the full dataset without reference material and blanks. Each dot represents the profile of one sample aliquot whereas lines delimit groups of samples that cluster together. **a)** PCA plots of the full dataset without reference material and blanks in positive mode. Components (Scr) 1 and 2 explain 9.79% and 7.59% of the variation respectively. **b)** PCA plots of the full dataset without reference material and blanks in negative mode. Components (Scr) 1 and 2 explain 13.20% and 10.30% of the variation respectively.

**a**

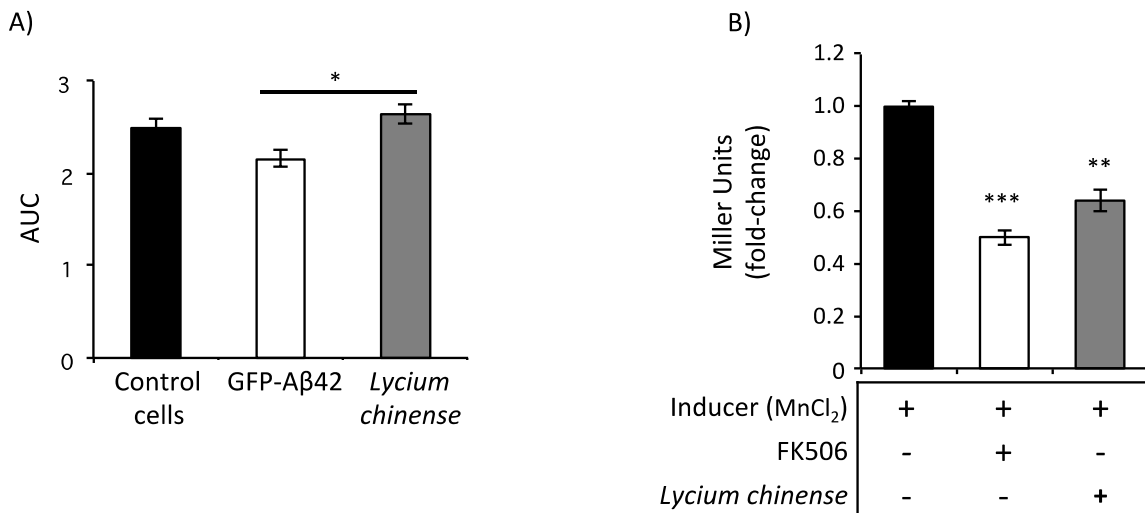


**b**



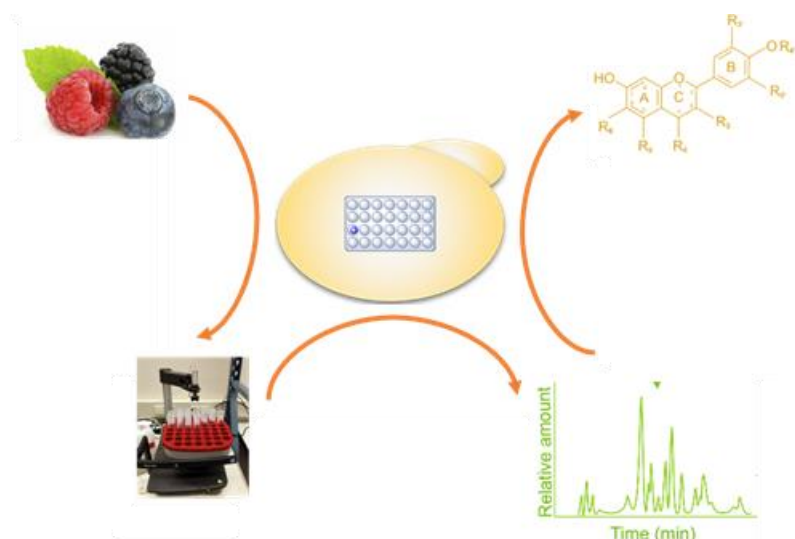
Yeast model of Crz1 activation

**Figure 3:** Schematic representation of yeast models. **a.** Yeast models of neurodegenerative diseases. Humanized yeast strains encoding chimeric fusions of disease genes with GFP (green fluorescent protein) under the control of a galactose-inducible (GAL1) promoter. Expression of A $\beta$ 42,  $\alpha$ Syn (SNCA gene), FUS and HTTQ103 leads to the formation of toxic aggregates, recapitulating the pathological processes of Alzheimer's disease (AD), Parkinson's disease (PD), Amyotrophic lateral sclerosis (ALS), and Huntington's disease (HD), respectively. **b.** Yeast model of Crz1 (NFAT orthologue) activation. Crz1 is activated by the serine/threonine protein phosphatase calcineurin (composed by CnA and CnB subunits). The reporter strain encodes the *lacZ* gene under the control of a promoter containing Crz1-binding sites, allowing the assessment of Crz1 activation through the measurement of  $\beta$ -galactosidase activity.

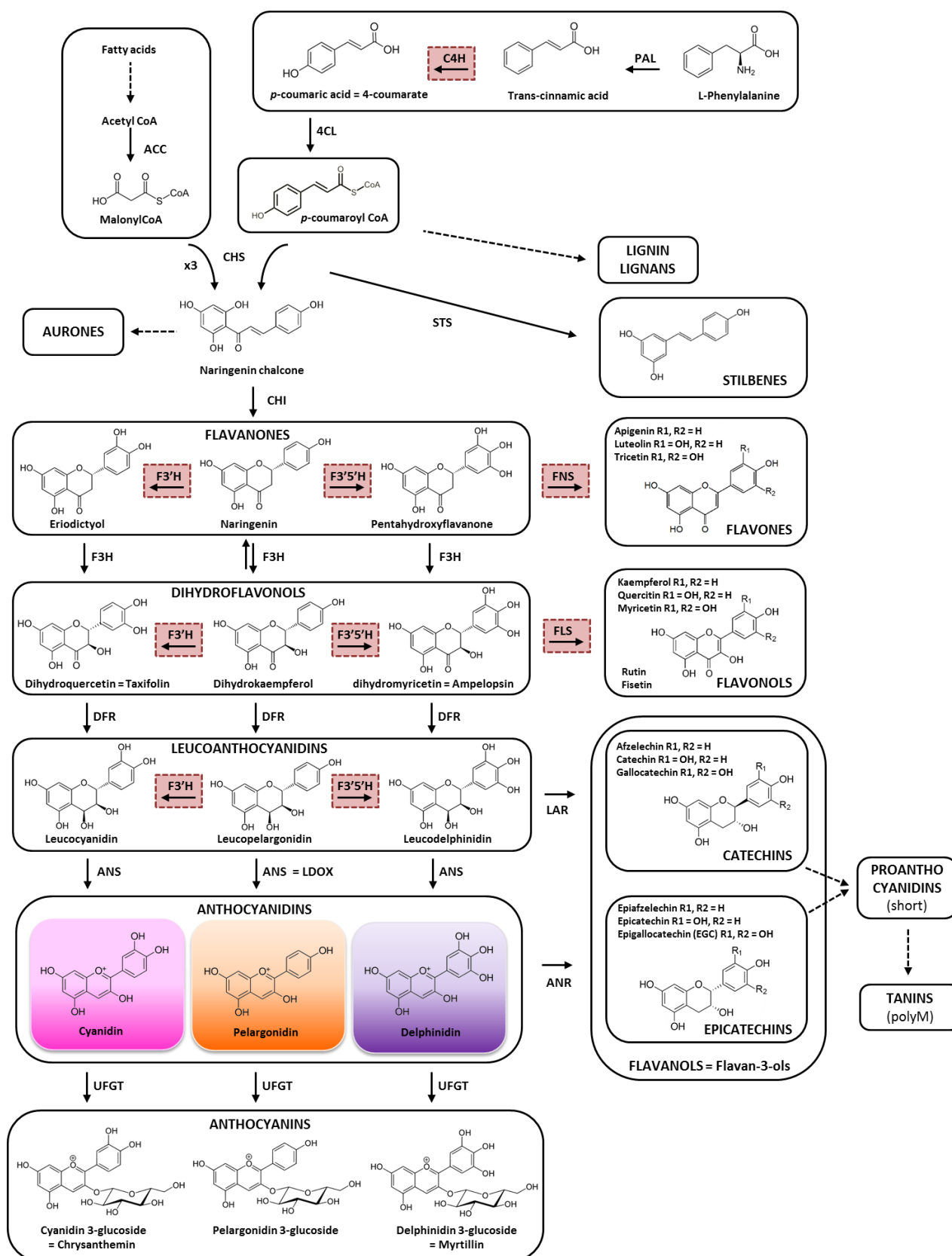


**Figure 4:** Identification of bioactivities in the berry germplasm. a) Strategy used for the discovery of bioactive compounds for Alzheimer's disease. BY4741\_erg6 recombinant yeast cells expressing GFP-Aβ42 were grown in SD galactose medium and cell viability was assessed by growth curves for 24 h in cells challenged or unchallenged with *Lycium chinense* polyphenol extracts. Cells containing the empty vector were used as controls. The areas under the curve (AUC) were integrated using the Origin software (OriginLab, Northampton, MA). Values represent the mean ± standard error of mean (SEM) of at least three biological replicates.

b) Strategy used for the discovery of bioactive compounds with anti-inflammatory potential. BY4741 yeast cells encoding CDRE<sub>promoter</sub>-lacZ were treated or not with *Lycium chinense* polyphenol extracts and induced with 1.8 mM MnCl<sub>2</sub>. Crz1 activation was assessed by monitoring β-galactosidase activity using ortho-nitrophenyl-β-galactoside (ONPG). The immunosuppressant FK506, a well-known inhibitor of calcineurin (Dumont 2000) was used as a positive control. The values represent the mean ± SEM of at least three biological replicates, \*p < 0.05, \*\*p < 0.01, \*\*\*p < 0.001.

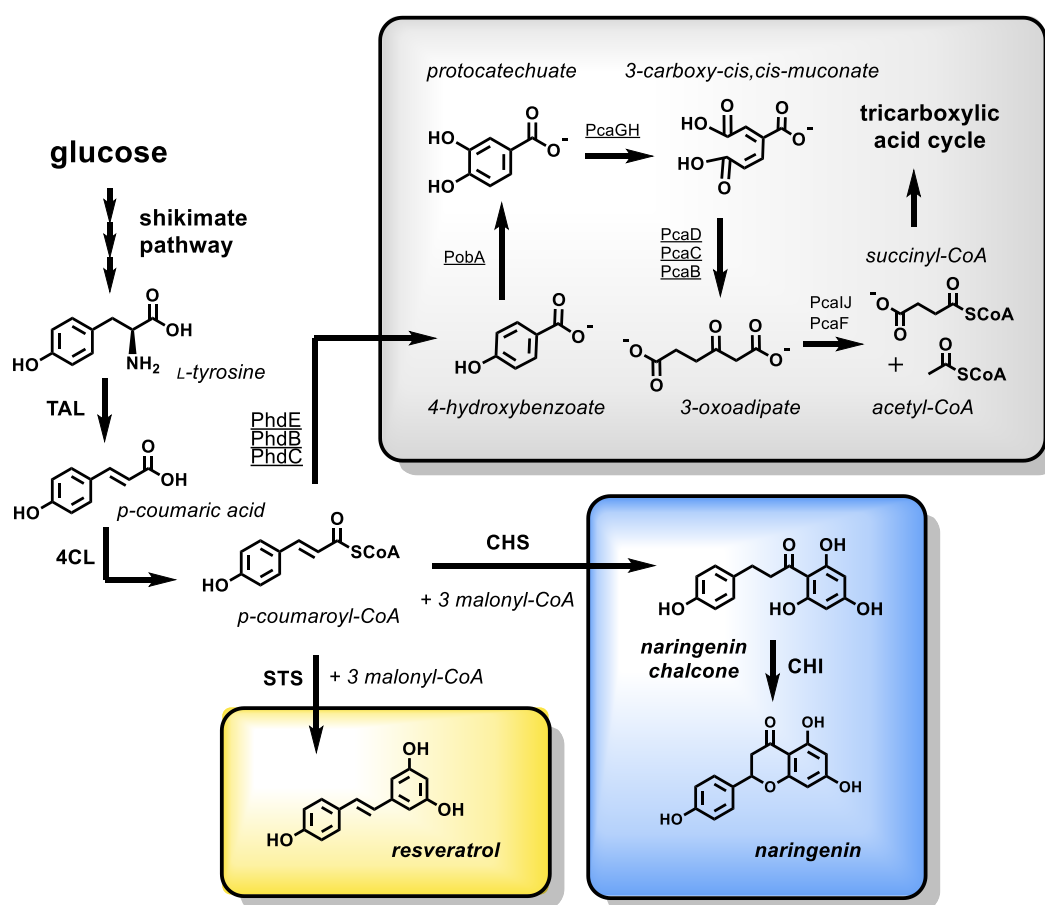


**Figure 5.** Yeast as a discovery platform of therapeutic compounds. Iterative bio-guided fractionation of complex mixtures allows the discovery of small molecules improving cell growth, by rescue of protein aggregates toxicity, or inhibiting inflammatory processes. High-throughput formats aligned with chemical identification by LC-MS approaches accelerate the identification of candidate compounds.



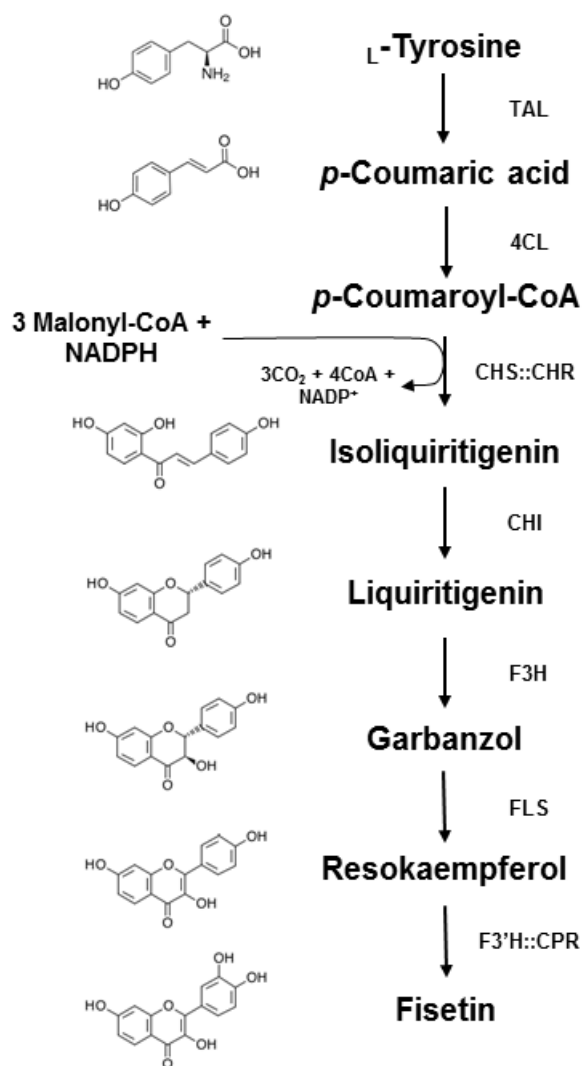
**Figure 6:** Polyphenol biosynthetic pathway in plants (modified from Falcone Ferreyra et al. 2012). PAL, phenylalanine ammonia-lyase; C4H, cinnamate 4-hydroxylase; 4CL, 4-coumaroyl-CoA ligase; CHS,

chalcone synthase; CHI, chalcone isomerase; F3H, flavanone 3-hydroxylase; F3'5'H, flavonoid 3'-5'-hydroxylase; DFR, dihydroflavonol 4-reductase; ANS, anthocyanidin synthase; ANR, anthocyanidin reductase; UFGT, flavonoid 3-*O*-glucosyltransferase. LAR, leucoanthocyanidin reductase; FNS, flavone synthase; FLS, flavonol synthase; STS, stilbene synthase. P450 enzymes are highlighted.

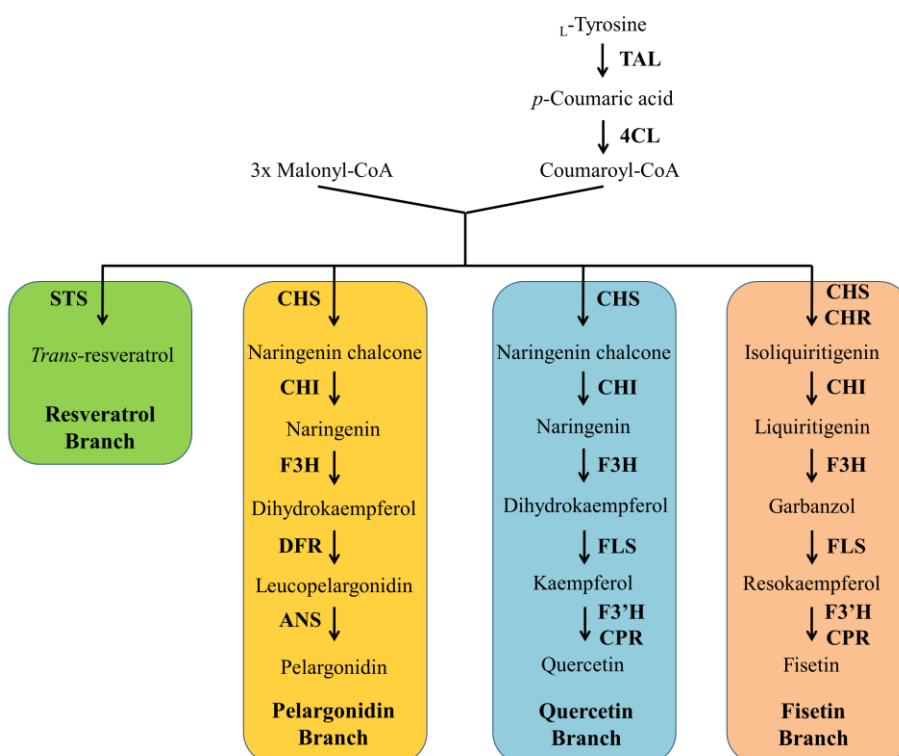


**Figure 7:** Overview of the endogenous phenylpropanoid degradation and the engineered pathway for polyphenol synthesis in *C. glutamicum*. One of the precursors for the production of the stilbene *trans*-resveratrol (catalyzed by stilbene synthase, STS) or for naringenin chalcone (catalyzed by chalcone synthases, CHS) is *p*-coumaroyl-CoA. Naringenin chalcone isomerizes to the (2*S*)-flavanone naringenin either spontaneously or catalyzed by the activity of chalcone isomerase (CHI). In *C. glutamicum*, *p*-coumaric acid can be degraded to 4-hydroxybenzoate by a CoA-dependent,  $\beta$ -oxidative deacetylation pathway. 4-Hydroxybenzoate is subsequently hydroxylated to protocatechuate, which is catabolized to succinyl-CoA and acetyl-CoA by the  $\beta$ -ketoacid pathway. Genes coding for the underlined enzymes were deleted in the course of the construction of the platform strain *C. glutamicum* DelAro<sup>4</sup>.

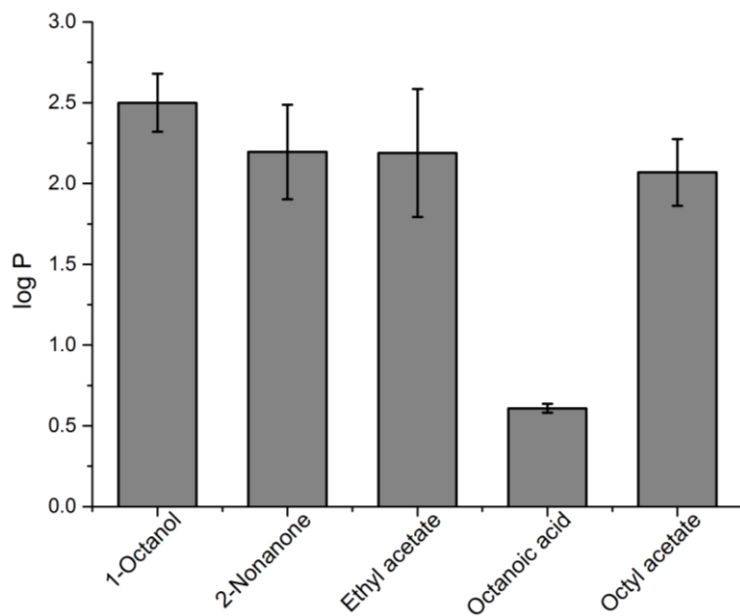




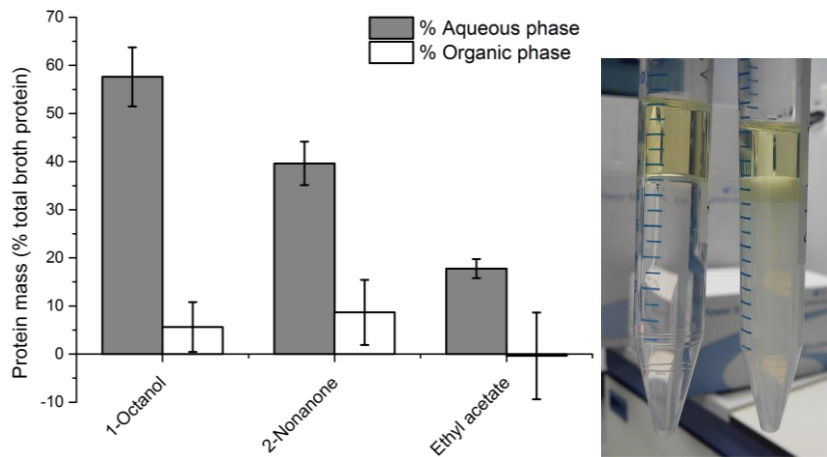
**Figure 8:** Heterologous biosynthetic pathway for fisetin assembled in *E. coli*. TAL, tyrosine ammonia-lyase; 4CL, 4-coumaroyl-CoA ligase; CHS::CHR, chalcone synthase::chalcone reductase fusion; CHI, chalcone isomerase; F3H, flavanone 3-hydroxylase; FLS, flavonol synthase; F3'H::CPR, flavonoid 3'-hydroxylase::cytochrome P450 reductase fusion.



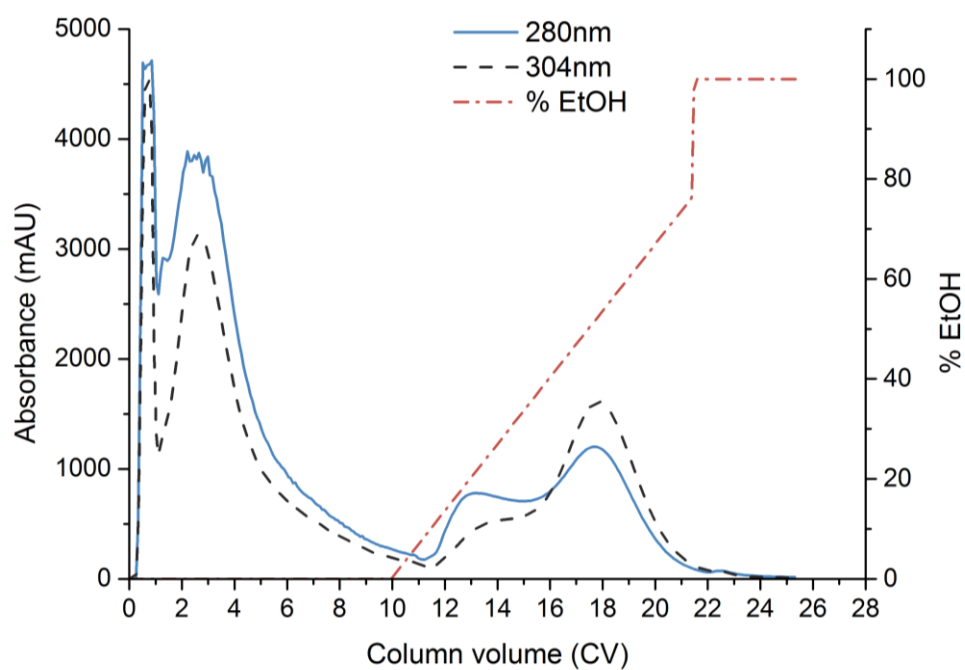
**Figure 9:** Heterologous biosynthetic pathways introduced into the genome-scale models of *C. glutamicum* and *L. lactis* leading to the four target compounds: i) *trans*-resveratrol ii) pelargonidin iii) quercetin and iv) fisetin. Two heterologous enzymes (TAL and 4CL) are common for the biosynthesis of the phenolic molecules. Enzyme abbreviations: TAL, tyrosine ammonia-lyase; 4CL, 4-coumaroyl-CoA ligase; STS, stilbene synthase; CHS, chalcone synthase; CHR, chalcone reductase; CHI, chalcone isomerase; F3H, flavanone 3-hydroxylase; FLS, flavonol synthase; F3'H, flavonoid 3' hydroxylase; CPR, cytochrome P450 reductase.



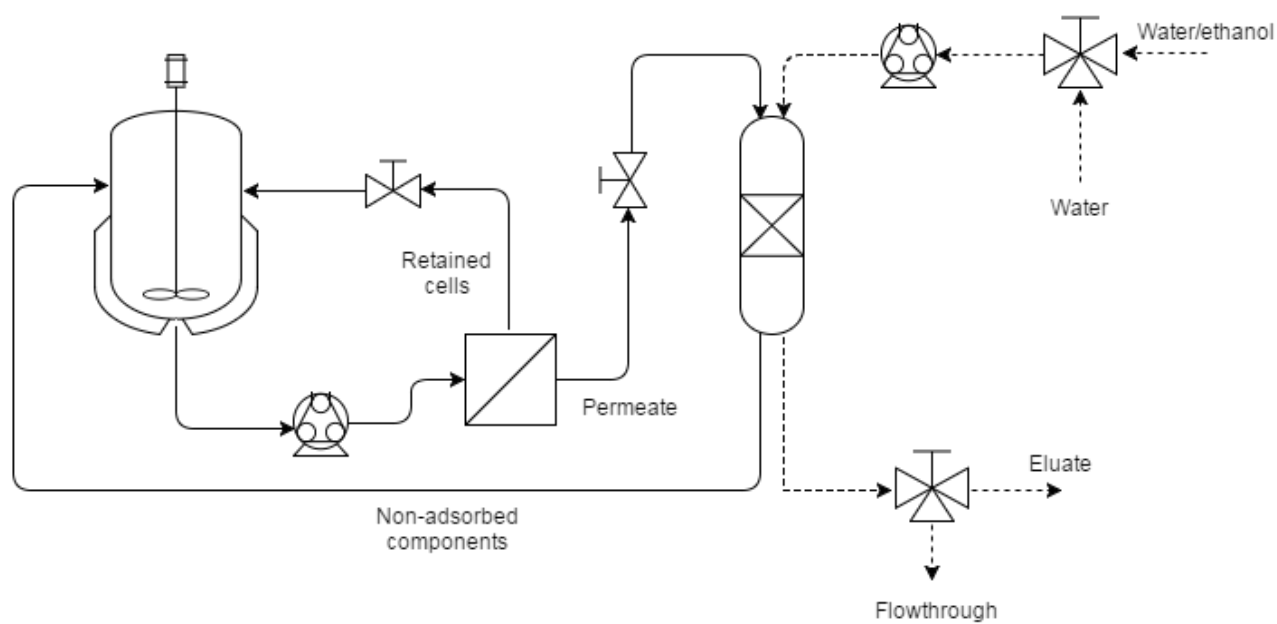
**Figure 10:** Partition coefficients in the form of log P values of *trans*-resveratrol using different organic solvents using *C. glutamicum* fermentation broth as the aqueous phase.



**Figure 11:** On the left, protein concentration (expressed as percentage of the initial broth protein concentration) in the organic and aqueous phases, after performing liquid extraction. The concentration in the organic phase is almost zero (considering the error bars), indicating that some protein stayed in the interphase. On the right, the formation of an interphase after the broth had been extracted with organic solvent (a mixture of heptanoic acid and hexyl acetate in this case) was clear.



**Figure 12:** Chromatogram following a pulse injection of 10 ml of *C. glutamicum* spent fermentation broth in a 30 ml column packed with Amberlite XAD-16 resin. A washing step was performed with 10 column volumes (CVs) of MilliQ water. Afterwards, elution was performed with a water/ethanol gradient for 12 CV, followed by a final regeneration step with 100% ethanol.



**Figure 13:** Experimental set-up for the “Continuous Product Removal” mode.

[Click here to view linked References](#)

## **SUPPLEMENTARY TABLES**

### **Title**

BacHBerry: BACterial Hosts for production of Bioactive phenolics from bERRY fruits

### **Running Title**

BacHBerry

### **Authors**

BacHBerry consortium

\*Corresponding author:

Alexey Dudnik

Email: [adud@biosustain.dtu.dk](mailto:adud@biosustain.dtu.dk)

Phone: + 45 93 51 11 01

Fax: +45 45 25 80 01

**Table S1** – Decorating glycosyl groups of anthocyanins found in different berry species. The presence of a glycosylation pattern is noted with + whereas the absence is noted with -.

	3-O-arabinoside	3-O-glucoside	3-O-galactoside	3-O-rutinoside	3-O-sambubioside	3-O-sophoroside	3-O-xyloside	3,5-diglucoside	3-O-glucosyl-rutinoside	3-O-sambubiosyl- 5-O-glucoside	3-O-xylosyl-rutinoside	3-O-galactoside-5-O-glucoside	3-O-gentibioside	3-O-rutinoside-5-O-glucoside
<i>Fragaria</i> <sup>1</sup>	+	+	-	+	-	-	-	-	-	-	-	-	-	-
<i>Ribes</i> <sup>1</sup>	-	+	+	+	+	+	-	-	+	-	+	-	-	-
<i>Rubus</i> <sup>1</sup>	-	+	-	+	-	+	+	+	+	-	+	-	-	-
<i>Vaccinium</i> <sup>1</sup>	+	+	+	-	-	-	-	-	-	-	-	-	-	-
<i>Vitis</i> <sup>1</sup>	-	+	-	-	-	-	-	-	-	-	-	-	-	-
<i>Lonicera</i> <sup>2</sup>	-	+	-	+	-	-	-	+	-	-	-	-	+	-
<i>Lycium</i> <sup>3</sup>	-	-	-	-	-	-	-	+	-	-	-	+	-	+
<i>Berberis</i> <sup>4,5</sup>	-	+	-	+	+	-	-	+	-	+	-	-	-	+
<i>Ugni</i> <sup>4</sup>	-	+	-	-	-	-	-	-	-	-	-	-	-	-

<sup>1</sup> Phenol explorer database (<http://phenol-explorer.eu>)

<sup>2</sup> Chaovanalikit et al., 2004

<sup>3</sup> Zheng et al., 2011

<sup>4</sup> Ruiz et al., 2010

<sup>5</sup> Ruiz et al., 2013

**Table S2** – Decorating acyl groups of anthocyanins found in a variety of berry species

	3-O-(6'',acetyl-glycoside)	3-O-(6'',malonyl-glycoside)	3-O-(6'',succinyl-glycoside)	3-O-(6'',dioxalyl-glycoside)	3-O-(6'',p-coumaryl-glycoside)	3-O-(feruloyl-glycoside)	3-O-(6'',caffeoyl-glycoside)
<i>Fragaria</i> <sup>1</sup>	-	+	+	-	-	-	-
<i>Ribes</i> <sup>1</sup>	-	-	-	-	+	-	+
<i>Rubus</i> <sup>1</sup>	-	+	-	+	-	-	-
<i>Vaccinium</i> <sup>1</sup>	+	-	-	-	-	-	-
<i>Vitis</i> <sup>1,2</sup>	+	-	-	-	+	-	+
<i>Lycium</i> <sup>3</sup>	-	-	-	-	+	+	+

<sup>1</sup> Phenol explorer database (<http://phenol-explorer.eu>)

<sup>2</sup> Figueiredo-Gonzalez et al. 2012

<sup>3</sup> Zheng et al. 2011



**Table S3** – List of all the species/varieties and harvest locations of the germplasm collections

Species	Variety/line	Harvest location
<i>Aristotelia chilensis</i> (Molina)Stuntz	-	Chile <sup>i</sup>
<i>Berberis buxifolia</i> Lam.	-	Chile <sup>ii</sup>
<i>Lonicera caerulea</i> L.	260-11	Russia <sup>iii</sup>
<i>Lonicera caerulea</i> L.	838-12	Russia <sup>iii</sup>
<i>Lonicera caerulea</i> L.	Avacha	Russia <sup>iii</sup>
<i>Lonicera caerulea</i> L.	Bazhovskaya	Russia <sup>iii</sup>
<i>Lonicera caerulea</i> L.	Leningradskii Velikan	Russia <sup>iii</sup>
<i>Lonicera caerulea</i> L.	Malvina	Russia <sup>iii</sup>
<i>Lonicera caerulea</i> L.	Morena	Russia <sup>iii</sup>
<i>Lonicera caerulea</i> L.	Nimfa	Russia <sup>iii</sup>
<i>Lonicera caerulea</i> L.	Pavlovskaya	Russia <sup>iii</sup>
<i>Lonicera caerulea</i> L.	S322-23	Russia <sup>iii</sup>
<i>Lonicera caerulea</i> L.	Solovey	Russia <sup>iii</sup>
<i>Lonicera caerulea</i> L.	Start	Russia <sup>iii</sup>
<i>Lonicera caerulea</i> L.	Suvenir	Russia <sup>iii</sup>
<i>Lycium chinense</i> Mill.	-	China <sup>iv</sup>
<i>Ribes</i> spp. L.	1126	United Kingdom <sup>v</sup>
<i>Ribes</i> spp. L.	1196	United Kingdom <sup>v</sup>
<i>Ribes</i> spp. L.	2174	United Kingdom <sup>v</sup>
<i>Ribes</i> spp. L.	2292-1	United Kingdom <sup>v</sup>
<i>Ribes</i> spp. L.	644217	United Kingdom <sup>v</sup>
<i>Ribes</i> spp. L.	-	United Kingdom <sup>v</sup>
<i>Ribes</i> spp. L.	-	Chile <sup>ii</sup>
<i>Ribes aureum</i> Pursh	R2-1-2	United Kingdom <sup>v</sup>
<i>Ribes bethmontii</i> Jancz.	XW6	United Kingdom <sup>v</sup>
<i>Ribes distans</i> Jancz.	2269	United Kingdom <sup>v</sup>
<i>Ribes divaricatum</i> Dougl.	Montana	United Kingdom <sup>v</sup>
<i>Ribes fuscescens</i> Jancz.	XB6	United Kingdom <sup>v</sup>
<i>Ribes futurum</i> Jancz.	XR4	United Kingdom <sup>v</sup>
<i>Ribes giraldii</i> Jancz.	B2-3-1	United Kingdom <sup>v</sup>
<i>Ribes grossularia</i> L.	1815-21	United Kingdom <sup>v</sup>
<i>Ribes grossularia</i> L.	1824-37	United Kingdom <sup>v</sup>
<i>Ribes grossularia</i> L.	9492-5	United Kingdom <sup>v</sup>
<i>Ribes grossularia</i> L.	96101-1	United Kingdom <sup>v</sup>
<i>Ribes grossularoides</i> Maxim.	H1-12-1	United Kingdom <sup>v</sup>
<i>Ribes holosericeum</i> Otto& A.Dietr.	XR2	United Kingdom <sup>v</sup>
<i>Ribes irriguum</i> Douglas	H1-5-1	United Kingdom <sup>v</sup>
<i>Ribes koehneanum</i> Jancz.	XR3	United Kingdom <sup>v</sup>
<i>Ribes nevadense</i> Kellog	-	United Kingdom <sup>v</sup>
<i>Ribes nigrum</i> L.	Ben Alder	United Kingdom <sup>v</sup>
<i>Ribes nigrum</i> subsp. <i>europaeum</i> x <i>R. nigrum</i> subsp. <i>sibiricum</i> x <i>R.dikuscha</i>	Almiai	Russia <sup>iii</sup>
<i>Ribes nigrum</i> subsp. <i>europaeum</i> x <i>R. nigrum</i> subsp. <i>sibiricum</i> x <i>R.dikuscha</i> x <i>R.nigrum</i> subsp. <i>scandinavicum</i>	Azhurnaya	Russia <sup>iii</sup>
<i>Ribes nigrum</i> subsp. <i>europaeum</i> x <i>R. nigrum</i> subsp. <i>sibiricum</i> x <i>R.dikuscha</i> x <i>R.nigrum</i> subsp. <i>scandinavicum</i>	Bolero	Russia <sup>iii</sup>

<i>Ribes nigrum</i> subsp. <i>europaeum</i> x <i>R. nigrum</i> subsp. <i>sibiricum</i>	Buraya Dal'nevostochnaya	Russia <sup>iii</sup>
<i>Ribes nigrum</i> subsp. <i>europaeum</i> x <i>R. nigrum</i> subsp. <i>sibiricum</i> x <i>R.dikuscha</i> x <i>R.nigrum</i> subsp. <i>scandinavicum</i>	Charovnitsa	Russia <sup>iii</sup>
<i>Ribes nigrum</i> subsp. <i>europaeum</i> x <i>R. nigrum</i> subsp. <i>sibiricum</i> x <i>R.dikuscha</i>	Dochka	Russia <sup>iii</sup>
<i>Ribes nigrum</i> subsp. <i>europaeum</i> x <i>R. nigrum</i> subsp. <i>sibiricum</i> x <i>R.dikuscha</i> x <i>R.nigrum</i> subsp. <i>scandinavicum</i>	Krasa L'vova	Russia <sup>iii</sup>
<i>R.nigrum</i> subsp. <i>scandinavicum</i> x <i>R.dikuscha</i>	Lazur'	Russia <sup>iii</sup>
<i>Ribes nigrum</i> subsp. <i>europaeum</i> x <i>R. nigrum</i> subsp. <i>sibiricum</i> x <i>R.dikuscha</i> x <i>R.nigrum</i> subsp. <i>scandinavicum</i>	Malen'kii Prints	Russia <sup>iii</sup>
<i>Ribes nigrum</i> subsp. <i>europaeum</i> x <i>R. nigrum</i> subsp. <i>sibiricum</i> x <i>R.dikuscha</i> x <i>R.nigrum</i> subsp. <i>scandinavicum</i>	Mila	Russia <sup>iii</sup>
<i>Ribes nigrum</i> subsp. <i>europaeum</i> x <i>R. nigrum</i> subsp. <i>sibiricum</i> x <i>R.dikuscha</i> x <i>R.nigrum</i> subsp. <i>scandinavicum</i>	Monisto	Russia <sup>iii</sup>
<i>Ribes nigrum</i> subsp. <i>europaeum</i> x <i>R. nigrum</i> subsp. <i>sibiricum</i> x <i>R.dikuscha</i> x <i>R.nigrum</i> subsp. <i>scandinavicum</i> x <i>R.janczewskii</i>	Muravushka	Russia <sup>iii</sup>
<i>Ribes nigrum</i> subsp. <i>europaeum</i> x <i>R. nigrum</i> subsp. <i>sibiricum</i> x <i>R.dikuscha</i> x <i>R.nigrum</i> subsp. <i>scandinavicum</i>	Ocharovanie	Russia <sup>iii</sup>
<i>Ribes nigrum</i> subsp. <i>europaeum</i> x <i>R. nigrum</i> subsp. <i>sibiricum</i> x <i>R.dikuscha</i> x <i>R.nigrum</i> subsp. <i>scandinavicum</i>	Orlovskaya serenada	Russia <sup>iii</sup>
<i>Ribes nigrum</i> subsp. <i>europaeum</i> x <i>R. nigrum</i> subsp. <i>sibiricum</i> x <i>R.dikuscha</i> x <i>R.nigrum</i> subsp. <i>scandinavicum</i>	Orlovskii Val's	Russia <sup>iii</sup>
<i>Ribes nigrum</i> subsp. <i>europaeum</i> x <i>R. nigrum</i> subsp. <i>sibiricum</i> x <i>R.dikuscha</i>	Sensei	Russia <sup>iii</sup>
<i>Ribes nigrum</i> subsp. <i>europaeum</i> x <i>R. nigrum</i> subsp. <i>sibiricum</i> x <i>R.dikuscha</i> x <i>R.nigrum</i> subsp. <i>scandinavicum</i>	Sharovidnaya	Russia <sup>iii</sup>
<i>Ribes nigrum</i> subsp. <i>europaeum</i> x <i>R. nigrum</i> subsp. <i>sibiricum</i> x <i>R.dikuscha</i>	Syuita Kievskaya	Russia <sup>iii</sup>
<i>Ribes nigrum</i> subsp. <i>europaeum</i> x <i>R. nigrum</i> subsp. <i>sibiricum</i> x <i>R.nigrum</i> subsp. <i>scandinavicum</i>	Ukrainka	Russia <sup>iii</sup>
<i>Ribes nigrum</i> subsp. <i>europaeum</i> x <i>R. nigrum</i> subsp. <i>sibiricum</i> x <i>R.dikuscha</i> x <i>R.nigrum</i> subsp. <i>scandinavicum</i>	Veloi	Russia <sup>iii</sup>
<i>Ribes nigrum</i> subsp. <i>europaeum</i> x <i>R. nigrum</i> subsp. <i>sibiricum</i> x <i>R.dikuscha</i> x <i>R.nigrum</i> subsp. <i>scandinavicum</i>	Vospominanie	Russia <sup>iii</sup>
<i>Ribes nigrum</i> subsp. <i>europaeum</i> x <i>R. nigrum</i> subsp. <i>sibiricum</i> x <i>R.dikuscha</i> x <i>R.nigrum</i> subsp. <i>scandinavicum</i>	Zelenaya Dymka	Russia <sup>iii</sup>
<i>Ribes nigrum</i> subsp. <i>europaeum</i> x <i>R. nigrum</i> subsp. <i>sibiricum</i> x <i>R.dikuscha</i>	Zhuravushka	Russia <sup>iii</sup>

<i>Ribes petraeum</i> Wulf.	R2-1-2	United Kingdom <sup>v</sup>
<i>Ribes rotundifolium</i> Michx.	H1-3-1	United Kingdom <sup>v</sup>
<i>Ribes rubrum</i> L.	Red Poll	United Kingdom <sup>v</sup>
<i>Ribes rubrum</i> L.	1518	United Kingdom <sup>v</sup>
<i>Ribes rubrum</i> L.	1169	United Kingdom <sup>v</sup>
<i>Ribes rubrum</i> L.	1196-2	United Kingdom <sup>v</sup>
<i>Ribes rubrum</i> L.	1757-2	United Kingdom <sup>v</sup>
<i>Ribes rubrum</i> L.	Red Wing	United Kingdom <sup>v</sup>
<i>Ribes rubrum</i> L.	S11-3-36	United Kingdom <sup>v</sup>
<i>Ribes sativum</i> Syme	R7-3-3	United Kingdom <sup>v</sup>
<i>Ribes x saundersii</i> Jancz.	XB4	United Kingdom <sup>v</sup>
<i>Ribes succirubrum</i> Zabel ex Jancz.	XW8	United Kingdom <sup>v</sup>
<i>Ribes warcewiczii</i> Jancz.	R7-3-3	United Kingdom <sup>v</sup>
<i>Ribes watsonianum</i> Koehne	-	United Kingdom <sup>v</sup>
<i>Rubus armeniacus</i> Focke [= <i>R. procerus</i> auct.])	Himalayan Giant	United Kingdom <sup>v</sup>
<i>Rubus bartonii</i> Newton syn. 'Ashton Cross'	Ashton Cross	United Kingdom <sup>v</sup>
<i>Rubus fruticosus</i> L. agg.	B13	United Kingdom <sup>v</sup>
<i>Rubus fruticosus</i> L. agg.	B14	United Kingdom <sup>v</sup>
<i>Rubus fruticosus</i> L. agg.	Fantasia	United Kingdom <sup>v</sup>
<i>Rubus fruticosus</i> L. agg.	Karaka Black	United Kingdom <sup>v</sup>
<i>Rubus geoides</i> Sm	-	Chile <sup>ii</sup>
<i>Rubus ursinus</i> x <i>R. armeniacus</i> x <i>R. idaeus</i>	Black Butte	United Kingdom <sup>v</sup>
<i>R. ursinus</i> x <i>R. ursinus</i> x <i>R. idaeus</i> x <i>R. ursinus</i> x <i>R. allegheniensis</i> x <i>R. argutus</i>	Kotata	United Kingdom <sup>v</sup>
<i>Rubus ursinus</i> x <i>R. ursinus</i> x <i>R. idaeus</i>	Ollalie	United Kingdom <sup>v</sup>
<i>Rubus idaeus</i> L.	Glen Ample	United Kingdom <sup>v</sup>
<i>Rubus idaeus</i> L.	Glen Fyne	United Kingdom <sup>v</sup>
<i>Rubus idaeus</i> L.	Malling Admiral	United Kingdom <sup>v</sup>
<i>Rubus idaeus</i> L.	Malmer Szedler	United Kingdom <sup>v</sup>
<i>Rubus idaeus</i> L.	Octavia	United Kingdom <sup>v</sup>
<i>Rubus idaeus</i> L.	Polana	United Kingdom <sup>v</sup>
<i>Rubus idaeus</i> L.	Prestige	United Kingdom <sup>v</sup>
<i>Rubus idaeus</i> L.	Tulameen	United Kingdom <sup>v</sup>
<i>Rubus loganobaccus</i> L.H.Bailey	Nectarberry	United Kingdom <sup>v</sup>
<i>Rubus loganobaccus</i> L.H.Bailey	Boysenberry	United Kingdom <sup>v</sup>
<i>Rubus loganobaccus</i> L.H.Bailey	Loganberry	United Kingdom <sup>v</sup>
<i>Rubus loganobaccus</i> L.H.Bailey	LY59-10	United Kingdom <sup>v</sup>
<i>Rubus loganobaccus</i> L.H.Bailey	Riwaka Choice	United Kingdom <sup>v</sup>
<i>Rubus loganobaccus</i> L.H.Bailey	Sunberry	United Kingdom <sup>v</sup>
<i>Rubus loganobaccus</i> L.H.Bailey	Tayberry	United Kingdom <sup>v</sup>
<i>Rubus occidentalis</i> L.	Black Hawk	United Kingdom <sup>v</sup>
<i>Rubus occidentalis</i> L.	Huron	United Kingdom <sup>v</sup>
<i>Rubus occidentalis</i> L.	Plum Farmer	United Kingdom <sup>v</sup>
<i>Rubus phoenicolasius</i> Maxim.	Wineberry	Portugal <sup>vi</sup>
<i>Rubus brigantinus</i> Samp.	-	Portugal <sup>vi</sup>
<i>Rubus genevieri</i> Boreau.	-	Portugal <sup>vi</sup>
<i>Rubus henriquesii</i> Samp.	-	Portugal <sup>vi</sup>
<i>Rubus hochstetterorum</i> Seub.	-	Portugal <sup>vi</sup>
<i>Rubus sampaioanus</i> Sudre ex Samp.	-	Portugal <sup>vi</sup>

<i>Rubus vagabundus</i> Samp.	-	Portugal <sup>vi</sup>
<i>Ugni molinae</i> Turcz.	Red Pearl	Chile <sup>vii</sup>
<i>Vaccinium angustifolium</i> Ait. ex Benth	Blomidon	China <sup>viii</sup>
<i>Vaccinium corymbosum</i> L.	Bluecrop	China <sup>ix</sup>
<i>Vaccinium darrowii</i> Camp	Misty	China <sup>ix</sup>
<i>Vaccinium uliginosum</i> L.	-	China <sup>viii</sup>
<i>Vaccinium virgatum</i> Ait.	Climax	China <sup>ix</sup>
<i>Vaccinium vitis-idaea</i> L.	-	China <sup>viii</sup>

---

<sup>i</sup> Sampling location: Coyhaique, Región de Aysén, Chile

<sup>ii</sup> Sampling location: Punta Arenas, Región de Magallanes, Chile

<sup>iii</sup> Vavilov Institute collection: Pavlovsk, St-Petersburg, Russia

<sup>iv</sup> Sampling location: Beijing botanical garden, Beijing, China

<sup>v</sup> The James Hutton Institute collection: Invergowrie, Scotland, UK

<sup>vi</sup> Propagation site: Fataga Experimental Field, Odemira, Portugal

<sup>vii</sup> Sampling location: Mataquito, Región del Maule, Chile

<sup>viii</sup> Sampling location: Tahe, Heilongjiang Province, China

<sup>ix</sup> Sampling location: Dalian, Liaoning Province, China

[Click here to view linked References](#)

## **SUPPLEMENTARY MATERIALS**

### **Title**

BacHBerry: BACterial Hosts for production of Bioactive phenolics from bERRY fruits

### **Running Title**

BacHBerry

### **Authors**

BacHBerry consortium

\*Corresponding author:

Alexey Dudnik

Email: [adud@biosustain.dtu.dk](mailto:adud@biosustain.dtu.dk)

Phone: + 45 93 51 11 01

Fax: +45 45 25 80 01

## S1: Materials and Methods

*Plant material and extraction procedure:* A total of 112 different cultivars and species from the *Aristotelia*, *Berberis*, *Lonicera*, *Lycium*, *Ribes*, *Rubus*, *Ugni* and *Vaccinium* genera grown in different locations (see Supplementary Table 1) were harvested manually in the field and kept under cool conditions until they were transferred to -20°C storage. Frozen fruit of each species/cultivar was weighed (approximately 50 g) into a solvent-proof blender containing 150 ml of pre-cooled 50ng ml<sup>-1</sup> Morin (Sigma-Aldrich, Gillingham, UK) solution prepared with 0.2% formic acid in methanol solution. Samples were then homogenized and subsequently filtered using Whatman filter paper grade 1. The resulting filtrate was aliquoted and solvent-dried using a speed-vac followed by lyophilization. Dried extracts were flushed with N<sub>2</sub> and stored at -20°C until analysis by LC-ToF-MS.

*Phenolic profile determination by LC-ToF-MS:* Dried extracts from each species/cultivar were resolubilized in triplicate using 2 ml of a 75% methanol solution with 0.1% formic acid. From these extracts, 500 µl were decanted into filter vials, sealed with 0.45 m PTFE-lined screwcap (Thomson Instrument Company, London, UK) and transferred into the autosampler. The analysis was achieved in positive and negative modes across 9 batches each, using an Agilent LC-ToF-MS system (Agilent Technologies, Cheshire, UK) consisting of a quaternary pump (Agilent 1260), a diode-array-detector (DAD) (Agilent 1260) a temperature control device (Agilent 1260) and a Thermostat (Agilent 1290) coupled to an Agilent 6224 time-of-flight (ToF) instrument. Five microliters of the sample were injected onto a 2x150 mm (4 µm) C18 column fitted with a C18 4 x 2 mm Security Guard™ cartridge (Phenomenex, Torrance, CA, USA). Sample and column temperatures were maintained at 4°C and 30°C, respectively. The samples were eluted at a flow rate of 0.3 ml min<sup>-1</sup> using two mobile phases (A: 0.1% Formic acid in dH<sub>2</sub>O; B: 0.1% formic acid in 50:50 dH<sub>2</sub>O: Acetonitrile) with the following gradient: 0 min 5%B; 4 min 5% B; 32.00 min 100%B; 34.00 min 100% B; 36.00 min 5% B; 40.00 min 5% B. For optimal electrospray ionization conditions the nebulizer pressure, drying gas temperature and drying gas were set to 45 psi, 350°C and 3 l min<sup>-1</sup>, respectively. In addition the diode array detection (DAD) was performed at 254, 280 and 520 nm. Morin levels (internal standard) were integrated in Agilent Mass Hunter Quan software (v. B.06.00) and all the samples with deviations larger than 10% relative to the dataset mean were reinjected. For all samples, three aliquots were analysed across three different analytical batches.

*Component detection, peak alignment and integration:* All chromatograms were processed identically using the Agilent Software Profinder v B.06.00 (Agilent Technologies, Cheshire, UK) which combines peak finding and integration algorithms for high-throughput data generation. For positive and negative mode data, the batch recursive molecular feature was used with peak extraction restricted to 2.1-38.00 mins of the chromatography and a peak threshold set at 15000 counts. The algorithm was set-up to include potential adducts of +H, +Na<sup>+</sup>, +K<sup>+</sup> and +NH<sub>4</sub><sup>+</sup> in positive mode and -

H and +Cl<sup>-</sup> in negative mode restricting the compounds to a maximum of one charge state. The compound ion count threshold was set at two or more ions, and for alignment purposes the RT window was set at 0.70% + 0.60 mins and the mass window was set at 25 ppm + 2 mDa. A post-processing filter to restrict analysis to compounds with more than 15000 counts and present in at least 3 of the files in at least one sample group (species/line). The find-by-ion options were set to limit the extracted ion chromatogram (EIC) to the expected retention time +/- 0.40 minutes. The “Agile” algorithm was used for the integration of EIC, with a gaussian smoothing of 9 points applied before the integration with a Gaussian width of 3 points. Additionally, peak filters were set at over 15000 counts and the chromatogram formats were set to centroid when available and otherwise profile. Spectrum was extracted at 10% of peak height and excluded if the spectra within the m/z range used was above 20% of saturation. Finally a post-processing filter was applied and compounds with less than 15000 counts or present in less 3 files in at least one sample group (species/line) were excluded.

*Statistical analyses:* Statistical analyses of all data were performed using GenStat for Windows, 16<sup>th</sup> Edition (VSN international Ltd., Hemel Hempstead, UK). A principal component analysis (PCA), based on the correlation matrix was applied to all the QC samples to ensure that the blank, reference samples, and berry samples were well separated (data not shown). Two separate analysis of the metabolomics dataset on the berry samples were carried out for the positive and negative dataset. A PCA, using the correlation matrix, was used to generate PCA plots of the first 4 principal components, which were subsequently used for selecting the species samples providing the most separation between samples.

*Yeast plasmids, strains and transformation:* p426\_GAL1pr-GFP-A $\beta$ 42 was generated by inserting the sequence GFP-A $\beta$ 42, obtained by the double digestion of p416\_GPDpr-GFP-A $\beta$ 42 (kindly provided by Ian Macraedie, RMIT University, AU) with *Bam*HI/*Sma*I, into the p426 vector. p425\_GAL1pr-GFP-A $\beta$ 42 was generated by subcloning the sequence GAL1pr-GFP-A $\beta$ 42 into the *Sac*I/*Hind*III restriction sites of p425 vector. *S. cerevisiae* strains used in this study are BY4742 *erg6 MATa his3 $\Delta$ 1 leu2 $\Delta$ 0 lys2 $\Delta$ 0 ura3 $\Delta$  YBR082c::kanMX4* (EUROSCARF) and BY4742\_*CDRE-lacZ MATa his3 leu2 lys2 ura3 aur1::AUR1-C-4xCDRE-lacZ* (Prescott et al, 2010).

*Yeast growth conditions and extract treatments:* Synthetic complete (SC) medium [0.67% (w/v) yeast nitrogen base without amino acids (YNB) (Difco) and 0.79 g l<sup>-1</sup> complete supplement mixture (CSM) (QBiogene)], containing 2% (w/v) glucose was used for growth of Crz1-reporter strain. Crz1 activation was induced with 1.8 mM MnCl<sub>2</sub> (Stathopoulos & Cyert). FK506 (Cayman Chemicals) was used as positive control. Synthetic dropout (SD)-LEU-URA media [0.67% (w/v) YNB and 0.54 g l<sup>-1</sup> 6-amino acid dropout CSM<sub>ADE-HIS-LEU-LYS-TRP-URA</sub> (QBiogene)], supplemented with standard concentrations of the required amino acids and containing 1% (w/v) raffinose, were used for growth

of the AD yeast model. The repression or induction of disease protein expression was carried out in medium containing glucose (control, disease-protein OFF) and galactose (disease-protein ON) at a final concentration of 2% (w/v).

A pre-inoculum was prepared in appropriated raffinose or glucose (only for Crz1-activation model) medium and cultures were incubated overnight at 30°C under orbital shaking. Cells were diluted in fresh medium and cultures were incubated under the same conditions until the optical density at 600 nm ( $OD_{600\text{ nm}}$ ) reached  $0.5 \pm 0.05$  (log growth phase). To ensure synchronized cells cultures, cell suspensions were further diluted according to the equation  $OD_i \times V_i = (OD_f / (2^{(t/gt)})) \times V_f$ , where  $OD_i$  = initial optical density of the culture,  $V_i$  = initial volume of culture,  $OD_f$  = final optical density of the culture,  $t$  = time (usually 16 h),  $gt$  = generation time of the strain and  $V_f$  = final volume of culture. Readings were performed in a 96-well microtiter plate using a Biotek Power Wave XS plate spectrophotometer.

*Growth assays:* Yeast cultures were diluted to  $OD_{600\text{ nm}} 0.12 \pm 0.012$  in fresh medium supplemented or not with the indicated concentrations of extracts in a 96-well microtiter plate. After 2 h incubation at 30°C, cultures were further diluted to  $OD_{600\text{ nm}} 0.03 \pm 0.003$  in medium containing glucose or galactose and supplemented or not with extracts. The cultures were then incubated at 30°C with shaking for 24 h or 48 h (for the AD model) and cellular growth was kinetically monitored hourly by measuring  $OD_{600\text{ nm}}$ . The areas under the curve (AUC) were integrated using the Origin software (OriginLab, Northampton, MA).

*$\beta$ -Galactosidase assays:* For quantitative measurements of  $\beta$ -galactosidase activity,  $OD_{600\text{ nm}}$  of cells cultures were recorded just before cell lysis. Cells were then incubated with Y-PER Yeast Protein Extraction Reagent (ThermoFisher Scientific) in 96-well microtiter plates for 20 min at 37°C, after which LacZ buffer containing 4 mg l<sup>-1</sup> ONPG was added and plates incubated at 30°C. The  $OD_{420\text{ nm}}$  and  $OD_{550\text{ nm}}$  was periodically monitored using a Biotek Power Wave XS Microplate Spectrophotometer until the development of the yellow color. Miller units were calculated as described previously (Garcia et al, 2016).

*Fractionation of bioactive fruit extracts:* The freeze-dried extract of selected berries (eg *Rubus genevieri*, and *Rubus idaeus*) were re-suspended in Milli-Q water, filtered and adsorbed onto a semi-preparation column packed with a C18 resin (PREP C18 55-105  $\mu\text{m}$  125Å, Waters Co. USA) using an AKTA Explorer (GE USA). 0.1% Formic Acid (Sigma, USA) in Milli-Q water was chosen as mobile phase A and Methanol (Chromasolv purity  $\geq 99.9\%$ , Sigma, USA) as mobile phase B. The approach was modified for the different berry extracts (*Rubus genevieri*, *Rubus idaeus*) to achieve proper fractionation, by tuning the hydrophobicity of the mobile phase (water-methanol). Modifying loading



and elution mobile phase composition and gradient profile served to generate a maximum number of well identifiable peak/fractions for further bio-activity testing (ranging from 12 to 28 fractions). Fractions were brought under vacuum using a rapid-vac (Labconco, USA) to remove methanol for further bioactivity testing.

## References

- Garcia G, Santos CN do, Menezes R (2016) High-Throughput Yeast-Based Reporter Assay to Identify Compounds with Anti-inflammatory Potential. In: Matthiesen R (ed) Proteostasis. Springer, New York, USA, pp 441–452
- Prescott TAK, Ariño J, Kite GC, Simmonds MSJ (2012) Inhibition of human calcineurin and yeast calcineurin-dependent gene expression by *Jasminum humile* leaf and root extracts. J Ethnopharmacol 140:293–297. doi: 10.1016/j.jep.2012.01.020